# First-Year EUMETSAT Fellowship Programme Research Report

## Towards a Hybrid Machine Learning Approach for Enhancing the Use of Satellite Data over Sea-Ice Surfaces

**Cristina González Flórez**[1], Fabrizio Baordo[1] and Stéphanie Guedj[2]

[1]Danish Meteorological Institute; [2]Norwegian Meteorological Institute

February 2026

# Contents

# List of Acronyms

**AMSR2** Advanced Microwave Scanning Radiometer 2.

**AMSU-A** Advanced Microwave Sounding Unit - A.

**AROME** Applications of Research to Operations at MEsoscale.

**ASIP** Automated Sea-Ice Products.

**BT** Brightness Temperature.

**DA** Data Assimilation.

**DMI** Danish Meteorological Institute.

**ECMWF** European Centre for Medium-Range Weather Forecasts.

**EUMETSAT** European Organisation for the Exploitation of Meteorological Satellites.

**FG** First-Guess.

**GPU** Graphics Processing Unit.

**IFS** Integrated Forecasting System.

**MARS** Meteorological Archival and Retrieval System.

**MetNO** Norwegian Meteorological Institute.

**METOP** Meteorological Operational satellite.

**ML** Machine Learning.

**NEMO** Nucleus for European Modelling of the Ocean.

**NOAA** National Oceanic and Atmospheric Administration.

**NWP** Numerical Weather Prediction.

**ODB** Observational DataBase.

**RMSE** Root Mean Square Error.

**RTTOV-SCATT** Radiative Transfer for TOVS Scattering.

**SAR** Synthetic Aperture Radar.

**SURFEM-Ocean** SURface Fast Emissivity Model for Ocean.

**TIROS** Television Infrared Observation Satellite.

**TOVS** TIROS Operational Vertical Sounder.

# 1 Introduction

Despite significant improvements in the coverage of meteorological, land-surface, and marine observations over recent years, the global observing system remains unevenly distributed (Fellous, 2008). In situ observations are much denser in populated and easily accessible regions, while there are persistent gaps over the open oceans, high-latitude polar regions, and other remote areas due to logistical and infrastructural constraints.

Consequently, satellite remote sensing observations play a key role in providing near-global coverage and continuous monitoring of essential variables (Yang et al., 2013). Traditionally, and still in current operational practice, the integration of in situ and satellite observations within atmospheric and coupled prediction systems relies on advanced Data Assimilation (DA) techniques. They provide a physically consistent representation of the system state, enhancing initial conditions and thereby improving forecast accuracy (Wang and Randriamampianina, 2021). These systems have historically relied on physically based radiative transfer models to relate satellite radiances to geophysical variables. However, the limited accuracy of such models over complex and heterogeneous surfaces, including sea-ice, constrains the effective use of surface-sensitive satellite observations.

Satellite instruments measure radiances influenced by multiple Earth system components, including atmospheric gases, clouds, aerosols, and the Earth's surface. Depending on the spectral frequency and the associated channel weighting functions, some channels are primarily sensitive to the upper atmosphere, while others are strongly influenced by surface emission. For surface-sensitive channels, both the atmospheric and surface states should therefore ideally be estimated simultaneously from these radiances, which requires the development of a coupled DA system across different Earth system components. This approach is inherently complex for several reasons (Geer, 2024a). First, many surface properties—such as emissivity, roughness, or, in the case of sea-ice, its temperature and moisture profile, snow cover, and microphysical characteristics like grain structure—are poorly known at the global scale due to the scarcity of direct and representative observations. Second, current models for radiative transfer in surface media, including sea-ice, soil, snow, and vegetation, exhibit limited accuracy and are often unable to fully reproduce observed radiances.

In recent years, there has been growing scientific, operational, and socio-economic interest in improving weather prediction in polar regions. This interest is driven by Arctic amplification, which results in faster-than-average warming at high latitudes and rapid changes in surface conditions such as sea-ice extent, with implications for both regional and global weather and climate patterns (Turner and Marshall, 2011; Naakka et al., 2025). In addition, increasing economic activity and the emergence of new maritime routes in polar regions heighten the demand for reliable weather and sea-ice forecasts to support safety and operational planning (Inoue, 2021). Although satellite observations provide valuable coverage over these regions, their full potential has not yet been fully exploited due to the limitations described above. As a result, many microwave observations that are strongly sensitive to surface conditions are still not assimilated over sea-ice.

To address these limitations, hybrid empirical–physical approaches have recently emerged, combining the physical consistency of DA-based radiative transfer models with the flexibility of Machine Learning (ML) to capture complex, nonlinear relationships in satellite

observations. These approaches provide a practical solution to a circular dependency: accurate global analyses of surface properties require reliable satellite observations, but fully exploiting these observations depends on accurate surface radiative transfer models. By embedding ML within a physically consistent DA framework, hybrid methods train simultaneously an empirical model and also works out what its inputs should be, effectively resolving this "chicken-and-egg" cycle.

In this context, the European Centre for Medium-Range Weather Forecasts (ECMWF) has extended its atmospheric DA system to simultaneously estimate the state of the surface, in particular sea-ice, through the development of a hybrid ML based observation operator (Geer, 2024b), which has been implemented operationally in the Integrated Forecasting System (IFS) cycle 49r1 (Geer, 2024a). This sea-ice observation operator has initially been trained using information from the Advanced Microwave Scanning Radiometer 2 (AMSR2) and later extended to handle additional conical-scanning sensors. It employs a neural network to model sea-ice emissivity, while the surface emissivity of the ocean and atmospheric radiative transfer rely on physical modelling. The sea-ice state parameters themselves are purely empirical, statistically representing the condition of sea-ice and its snow cover as it affects surface emissivity. These parameters are not known a priori, lack direct physical interpretation, and must be inferred from observations. Consequently, the model can only be applied within a physically consistent retrieval or DA system.

Building on the latest ECMWF developments, during the first year of my European Organisation for the Exploitation of Meteorological Satellites (EUMETSAT) fellowship, we have extended this hybrid ML framework to the Advanced Microwave Sounding Unit - A (AMSU-A), a cross-track scanning sensor. For this purpose, data extracted from the ECMWF IFS 49r1, including observations from the AMSU-A instrument onboard the Meteorological Operational satellite (METOP)-B, have been used. This extension represents a significant step forward in addressing future observational requirements, particularly in the context of the upcoming EUMETSAT Polar System – Sterna constellation programme of small satellites.

Section 2 describes the basis of the hybrid ML approach, including the architecture of the network, how the layers are connected, the integration of physical modelling, and the introduction of angular dependency to address cross-track sensors specificities. It also details the construction of the dataset and the experimental setup. Section 3 presents the results of these experiments, and Section 4 summarizes the main findings. Finally, Section 6 outlines the planned next steps for the following year of the fellowship, taking into account a change in research focus motivated by the departure of my main advisor at Danish Meteorological Institute (DMI).

# 2 Method

The hybrid ML model is designed to jointly estimate daily maps of sea-ice concentration and physical sea-ice state, alongside a new sea-ice surface emissivity model, which together produce simulated Brightness Temperature (BT)s that optimally match AMSU-A observations over a full year. This study extends the framework from conical-scanning to cross-track sensors, enabling potential application to other satellite platforms.

## 2.1 Hybrid ML approach

In a hybrid empirical–physical modelling framework, the forward model consists of a combination of physically based components and data-driven elements. Selected parts of the system, in this specific case sea-ice surface emissivity, are represented using trainable ML models (e.g. neural networks), while other components are governed by fixed physical equations. These elements are coupled through physically consistent relationships to produce simulated observables. The simulated observables are then compared with actual measurements, in our case, AMSU-A radiances expressed as equivalent black-body BTs. The differences between simulated and observed quantities, also referred to as First-Guess (FG) departures to observations, are quantified through a loss function. This loss function, commonly used in a ML context, plays a role analogous to the cost function in a DA context. It is typically composed of several loss terms, including a primary misfit term, and may incorporate additional regularization or constraint terms to stabilize the training and enforce physically plausible solutions. The minimization of this loss function allows for the simultaneous optimization of the ML parameters and selected parameters of the physically based components, and, where relevant, of the system state. Larger values of the loss function indicate poorer agreement between simulations and observations, while smaller values indicate improved model performance.

Figure 1 provides a simplified overview of the training workflow, highlighting the main variables and terms involved in the hybrid ML model used in this study. Figure 2 offers a more detailed view of the model architecture, showing each layer along with its inputs (blue boxes) and outputs (purple boxes), the trainable parameters (indicated by black dotted outlines), and the associated loss terms. Most input variable names in Figures 1 and 2 are taken directly from the original datasets, with their meanings and details summarized in Table 2. Some inputs, however, are derived from those variables, and are specified in Table 3. The following paragraphs provide a detailed description of the role of each layer shown in Figure 2, and the full names of their outputs are summarized in Table 1.

The aim of this hybrid ML model is to jointly estimate daily maps of sea-ice concentration ($c_{\text{ice}}$) and sea-ice physical state ($X_{\text{ice\_obs}}$, $X_{\text{ice\_grid}}$), as well as a new model for sea-ice surface emissivity ($e_{\text{ice}}$), which together generate simulated BTs that best match AMSU-A observations over an entire year. The model is implemented in Keras, part of Tensorflow, using Python 3.10. The sea-ice physical state is not described explicitly, but through a statistical representation referred to as the empirical state, corresponding to a latent space in unsupervised ML terminology. The variables defining this state have no direct physical interpretation. Instead, their meaning emerges implicitly during the training of the sea-ice surface emissivity model, which plays the role of a decoder mapping the latent state to physically observable quantities. The lack of a direct physical meaning for the empirical-state variables is the compromise required when the surface state cannot be adequately constrained owing to the lack of globally valid measurements of sea-ice and snow properties and/or comprehensive physical models. The purpose of the empirical state is to encode variations in surface conditions that influence the microwave surface emissivity. No prior information on sea-ice concentration or outputs from ocean or sea-ice models is used during training, in order to avoid constraining the learned emissivity model with potentially biased or inconsistent existing knowledge.

The observed AMSU-A radiances are influenced by atmospheric absorption and scattering by gases and clouds, as well as by surface emission, reflection, and scattering. In the hybrid ML framework adopted here, the atmospheric radiative transfer is treated entirely physically. These terms are computed prior to training using 12h forecasts from the ECMWF IFS DA system through the Radiative Transfer for TOVS Scattering (RTTOV-SCATT) model (Bauer et al., 2006; Saunders et al., 2018). Importantly, the ECMWF IFS DA system implements all-sky assimilation, whereby radiances are assimilated directly regardless of whether the observed scene is clear, cloudy, or precipitating. This is possible because both the forecast and radiative transfer models are capable of accurately representing clouds and precipitation (Baordo and Geer, 2016; Geer et al., 2018).

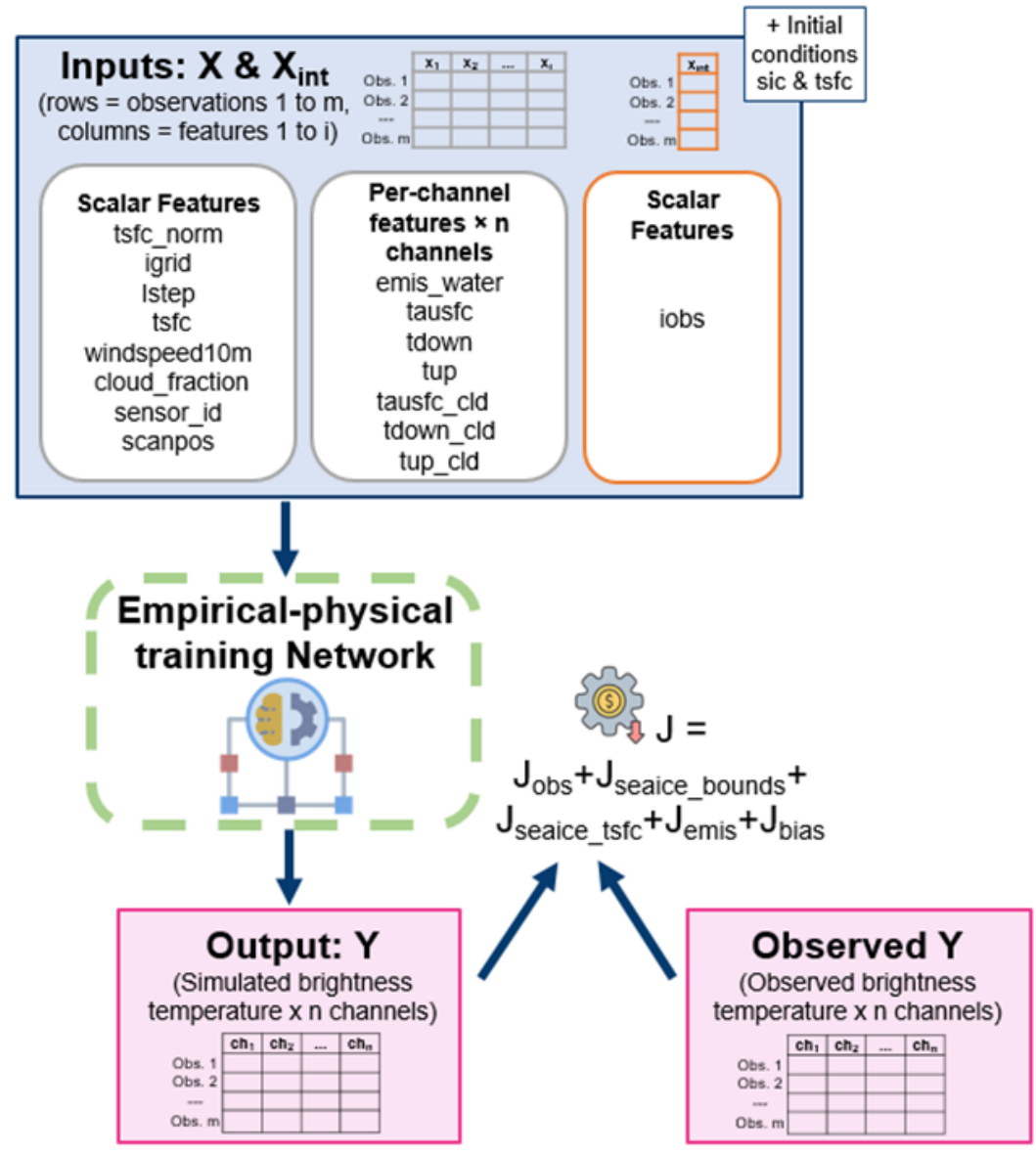


Figure 1: Training workflow of the hybrid machine learning model. Most input variables are taken directly from the original datasets (summarized in Table 2), while a few are derived from these variables (listed in Table 3).

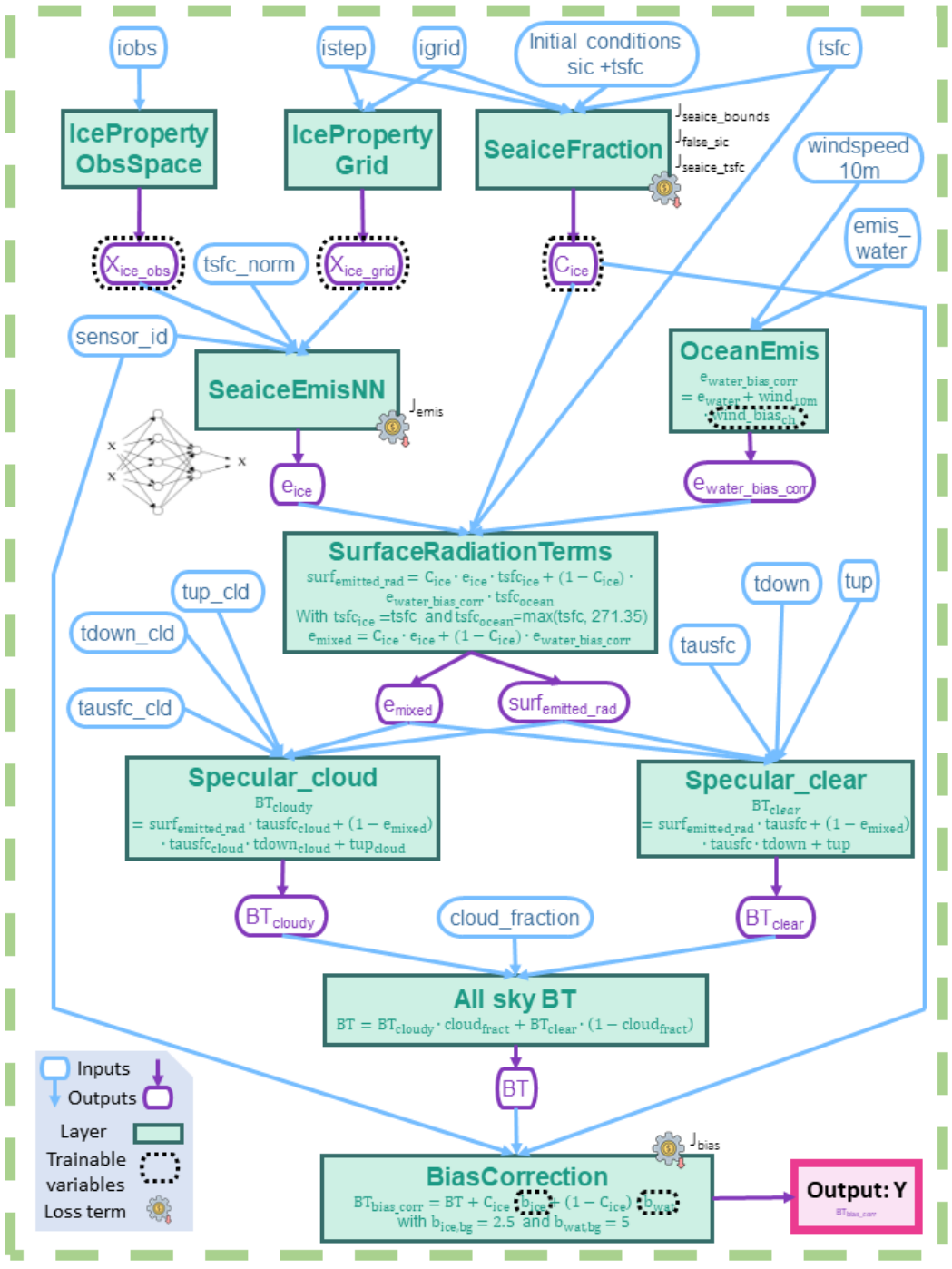


Figure 2: Detailed architecture of the hybrid machine learning network, showing each layer along with its inputs (blue boxes) and outputs (purple boxes), trainable parameters (indicated by dotted outlines), and associated loss terms. Most input variables are taken directly from the original datasets (summarized in Table 2), while a few are derived from these variables (listed in Table 3). Full descriptions of the outputs are provided in Table 1.

Table 1: Outputs of the layers shown in Figure 2 and their meanings.

| **Output** | **Description** | **Layer** |
|---|---|---|
| $X_{\text{ice\_obs}}$ | Empirical micro- and macro-physical property in the observation space. | IcePropertyObs Space |
| $X_{\text{ice\_grid}}$ | Empirical micro- and macro-physical properties in the grid space. | IcePropertyGrid |
| $C_{\text{ice}}$ | Daily map of sea-ice concentration in the grid space. | SeaiceFraction |
| $e_{\text{ice}}$ | Sea-ice surface emissivity. | SeaiceEmisNN |
| $e_{\text{water\_bias\_corr}}$ | Bias-corrected surface ocean emissivity. | OceanEmis |
| $e_{\text{mixed}}$ | Mixed surface emissivity combining sea-ice and ocean contributions weighted by sea-ice concentration. | SurfaceRadiation Terms |
| $\text{surf}_{\text{emitted\_rad}}$ | Mixed surface emitted radiation, computed as a weighted combination of sea-ice and ocean contributions using sea-ice concentration, respective surface emissivities, and surface skin temperature (with ocean temperatures limited to a minimum of 271.35K). | SurfaceRadiation Terms |
| $\text{BT}_{\text{cloudy}}$ | Simulated brightness temperature under cloudy conditions. | Specular_cloud |
| $\text{BT}_{\text{clear}}$ | Simulated brightness temperature under clear-sky conditions. | Specular_clear |
| BT | Final brightness temperature obtained by combining $\text{BT}_{\text{cloudy}}$ and $\text{BT}_{\text{clear}}$ using the effective cloud fraction. | All Sky BT |
| $\text{BT}_{\text{bias\_corr}}$ | Final bias corrected brightness temperature. | BiasCorrection |

Within the hybrid ML approach, the atmosphere is modelled using two independent sub-columns: one clear-sky and one cloudy-sky. The clear-sky sub-column simulates surface–atmosphere radiative interactions, including gaseous absorption, primarily by water vapour, and is implemented in the `Specular_clear` layer (see Figure 2). The cloudy-sky sub-column additionally accounts for the effects of clouds and precipitation using a delta–Eddington scattering solver, implemented in the `Specular_cloud` layer. In both cases, the surface is represented using a specular reflection approximation, in which the angular behaviour of the emitted and reflected radiation is governed by Fresnel equations at the surface boundary. This assumption is generally appropriate over open ocean and snow-free ice, where the surface behaves approximately as a smooth interface. However, over snow-covered ice or regions with complex microstructure, a more diffuse (Lambertian-like) behaviour—characterised by a weaker dependence of emitted radiance on the zenith (viewing) angle—may provide a more realistic representation and is currently under investigation (Bormann, 2022).

In the present implementation, the surface emits radiation proportional to the surface skin temperature multiplied by an effective surface emissivity $e$, and reflects downwelling radiation according to a reflectivity $1 - e$. Hence, the surface emissivity is treated as an effective parameter, capturing unresolved surface heterogeneity and modelling approximations rather than representing a purely physical material property. This simplification is necessary because microwave radiation penetrates from centimeters to meters into snow and ice, whereas the modelled surface corresponds only to the skin layer. Since the skin temperature is typically colder than the effective emitting temperature of the snow and ice, the effective emissivity can exceed 1, reducing the size of the reflectivity and, consequently, the amount of downwelling radiation reflected by the surface.

In this study, the surface consists of a mixture of sea-ice and open ocean. The effective mixed surface emissivity, computed in the `SurfaceRadiationTerms` layer, is therefore modelled as a linear combination of sea-ice and ocean emissivities, weighted by the sea-ice concentration within the satellite field of view.

Sea-ice concentration is defined here as the fraction of the ocean surface covered by ice or snow, excluding melt ponds but including icebergs and ice shelves. As described above, it is one of the variables estimated by the hybrid model and produced as daily maps on a fixed grid. To speed up the training, the sea-ice concentration maps are initialized using a monthly mean sea-ice concentration derived from the existing ECMWF IFS analysis. This field is produced by a dynamic–thermodynamic sea-ice model (LIM3, the ocean–ice component of Nucleus for European Modelling of the Ocean (NEMO); Rousset et al., 2015; Madec and the NEMO System Team, 2023) coupled to the ECMWF IFS atmospheric model.

The mixed surface emissivity must be estimated within the network in order to provide physically consistent boundary conditions for the atmospheric radiative transfer model. The ocean surface emissivity is treated as a known quantity and is taken from the ECMWF IFS DA system (SURface Fast Emissivity Model for Ocean (SURFEM-Ocean), Kilic et al., 2023), with an additional wind-speed- and channel-dependent bias correction applied in the `OceanEmis` layer. In contrast, the sea-ice surface emissivity is modelled using a two-layer neural network with a sigmoid activation function, implemented in the `SeaiceEmisNN` layer. This differs from Geer (2024a), where a single neural network was used. It is important to remark that this two-layer neural network is intended to cover all frequencies used from AMSU-A, all locations and all seasons.

The contrast in the surface emissivity between ocean and sea-ice is strongest and easiest to detect at microwave wavelengths, and is fundamentally what allows the inference of the sea-ice concentration from satellite observations. The final all-sky BT is obtained by combining the clear- and cloudy-sky sub-columns according to the effective cloud fraction in the `All sky BT` layer. A subsequent bias correction that is a function of the sea-ice concentration and per channel for sea-ice and for ocean water is applied in the `BiasCorrection` layer.

This bias corrected BT represents the final output of the hybrid model and constitutes the quantity that is directly compared with the observed AMSU-A BT during training. The training is performed iteratively over multiple epochs, where one epoch corresponds to a full pass through the entire training dataset. For computational efficiency, each

epoch is divided into smaller subsets of data, referred to as batches. For each batch, the simulated bias corrected BT is evaluated against the corresponding observations, and their differences are used to construct the loss function that drives the optimisation. Through backpropagation, the gradients of this loss are propagated through all trainable components of the network, enabling the joint adjustment of the sea-ice concentration, the empirical sea-ice state, the sea-ice emissivity neural network, and the bias correction parameters, while the physically based radiative transfer components remain fixed. In this way, the training process enforces consistency between simulated and observed radiances while preserving the physical constraints imposed by the hybrid model.

Equation 1 represents the loss function $J$ used during training, where $J_{obs}$ is the primary loss and the remaining terms correspond to additional regularization terms:

$$J = J_{obs} + J_{seaice\_bounds} + J_{false\_sic} + J_{seaice\_tsfc} + J_{bias} + J_{emis} \tag{1}$$

The main loss $J_{obs}$ measures the discrepancy between the AMSU-A BT observations ($BT^{\text{obs}}$) and the bias corrected simulated BT ($BT^{\text{sim}}_{bias\_corr}$):

$$J_{\text{obs}} = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m}\frac{\left(BT^{\text{obs}}_{ij} - BT^{\text{sim}}_{bias\_corr.ij}\right)^2}{r_j^2} \tag{2}$$

with the observation index $i$ over all observations $n$, channel index $j$ and $r$ the observation error.

The sea-ice concentration is constrained by $J_{seaice\_bounds}$, $J_{false\_sic}$ and $J_{seaice\_tsfc}$, which are computed in the `SeaiceFraction` layer. The first term enforces the physically admissible bounds of sea-ice concentration by the quadratic penalty:

$$J_{\text{seaice\_bounds}} = \frac{1}{n}\sum_{a,b}\frac{\max(C_{ab} - 1.0, 0.0)^2}{c^2_{\text{bgerr}}} \; + \; \frac{1}{n}\sum_{a,b}\frac{\min(C_{ab}, 0.0)^2}{c^2_{\text{bgerr}}} \tag{3}$$

where $C_{ab}$ denotes the sea-ice concentration at time index $a$ and spatial location $b$, and $c_{\text{bgerr}}$ is the background error. This term penalizes any values of $C_{ab}$ that fall outside the physically meaningful interval $[0, 1]$.

The second regularization term, $J_{\text{false\_sic}}$, penalizes the occurrence of spurious or trace sea-ice in regions where ice is climatologically unlikely. It is defined by a piecewise function that applies a quadratic penalty near the lower and upper trace-ice thresholds and a smoothly varying penalty in between:

$$J_{\text{false\_sic}} = \sum_d \begin{cases} \dfrac{(C_d - C_{\min})^2}{\sigma_{\text{false}}^2}, & C_{\min} < C_d \le C_{\min} + b, \\ H - \dfrac{s_{\text{peak}}}{\sigma_{\text{false}}^2} (C_{\text{mid}} - C_d)^2, & C_{\min} + b < C_d < C_{\max} - b, \\ \dfrac{(C_{\max} - C_d)^2}{\sigma_{\text{false}}^2}, & C_{\max} - b \le C_d < C_{\max}, \\ 0, & \text{otherwise}, \end{cases} \tag{4}$$

where $C_d$ is the sea-ice concentration at grid point $d$. The parameters $C_{\min}$ and $C_{\max}$ define the lower and upper bounds for trace ice concentrations, $C_{\text{mid}} = \frac{C_{\min}+C_{\max}}{2}$ is the midpoint between them, and $b = \frac{s_{\text{peak}}}{1+s_{\text{peak}}}(C_{\text{mid}} - C_{\min})$ is the width of the transition region, determined by the peak scaling $s_{\text{peak}}$. The background error for false sea-ice is $\sigma_{\text{false}}$, and $H = \frac{b\,(C_{\text{mid}} - C_{\min})}{\sigma_{\text{false}}^2}$ is chosen to ensure continuity of the penalty function.

Finally, the third term, $J_{\text{seaice\_tsfc}}$, penalizes the presence of sea-ice in regions with anomalously high surface skin temperatures, thereby reducing the risk of misattributing cloud or warm-surface signatures to sea-ice:

$$J_{\text{seaice\_tsfc}} = \frac{1}{n} \sum_{d\,:\,C_d > 0.01} 4.0 \, \max(T_{S,d} - 273.2, 0)\,, \tag{5}$$

where $T_{S,d}$ is the surface skin temperature at grid point $d$, and the sum is restricted to locations where the sea-ice concentration exceeds 0.01.

The loss function $J$ also includes a bias-correction term, $J_{\text{bias}}$, computed in the `Bias Correction` layer, which constrains the per-channel bias parameters for sea-ice and open ocean relative to their background values:

$$J_{\text{bias}} = \frac{1}{n} \sum_{j=1}^{m} \frac{(b_{\text{ice},j} - b_{\text{ice,bg}})^2 + (b_{\text{wat},j} - b_{\text{wat,bg}})^2}{b_{\text{bgerr}}^2} \tag{6}$$

Here, $b_{\text{ice},j}$ and $b_{\text{wat},j}$ denote the bias correction parameters for sea-ice and open ocean, respectively, for channel $j$, while $b_{\text{ice,bg}}$ and $b_{\text{wat,bg}}$ are their corresponding background (prior) values. The parameter $b_{\text{bgerr}}$ represents the background error that controls the strength of the constraint.

Finally, the last term, $J_{\text{emis}}$, computed in the `SeaiceEmisNN` layer, enforces a physical lower bound on the sea-ice emissivity for the lowest-frequency channel, in the case of AMSU-A, it is 23.8 GHz:

$$J_{\text{emis}}^{(c)} = \frac{1}{n} \sum_{i=1}^{n} \frac{\max\Big(-\big(e_i^{(c)} - e_{\text{bg}}\big),\, 0\Big)^2}{\sigma_{\text{bg}}^2} \tag{7}$$

where $e_i^{(c)}$ is the emissivity predicted by the network for observation $i$ at 23.8 GHz, $e_{\text{bg}}$ is

the background (minimum physically plausible) emissivity for this channel, $\sigma_{\mathrm{bg}}$ controls how sharply the penalty is applied, and $n$ is the number of observations.

It is important to note that this network is not intended to predict observations for other years, as the spatial distribution and physical properties of sea-ice evolve from year to year. Consequently, the conventional separation into training and test datasets, as commonly used in supervised ML applications, is not applicable. Instead, model performance is assessed based on its ability to fit the observations and to provide physically consistent estimates of sea-ice concentration.

## 2.2 Extension to cross-track scanning sensors

In this study, the hybrid ML framework originally developed for the conical-scanning sensor AMSR2 (Geer, 2024a,b) is extended to AMSU-A, a cross-track scanning sensor. Unlike conical-scanning instruments, which observe the surface at a fixed viewing geometry, cross-track sensors scan the Earth across the swath, perpendicular to the satellite ground track. Consequently, the viewing geometry of cross-track sensors varies systematically across the swath.

Microwave radiometers measure radiation in channels with nominal vertical (V) or horizontal (H) polarization, defined in the instrument reference frame. For conical scanners, the fixed viewing geometry ensures that this nominal polarization coincides with the actual polarization state of the radiation from the Earth. In contrast, for cross-track sensors, the rotation of the scan mirror induces a rotation of the polarization basis relative to the Earth-fixed frame, so that the signal received by the instrument is a scan-angle-dependent linear combination of the Earth-relative vertical and horizontal surface polarizations, even for channels that are nominally pure V or H.

To simulate the signal received by the instrument, the hybrid ML framework predicts the intrinsic Earth-relative sea-ice surface emissivity independently of the viewing geometry. This is achieved using a two-layer neural network, whose output consists of the vertical emissivity, $E_{\mathrm{V}}$, and horizontal emissivity, $E_{\mathrm{H}}$. These represent the physical surface properties in the Earth reference frame, before accounting for the instrument-specific polarization effects.

For conical scanners, the effective emissivity used in the radiative transfer corresponds directly to either $E_{\mathrm{V}}$ or $E_{\mathrm{H}}$, depending on the nominal channel polarization. For cross-track scanners, however, the effective emissivity seen by the instrument is obtained by applying a deterministic, scan-angle-dependent polarization mixing based on the known instrument geometry, described by:

$$E_{\mathrm{QV}} = E_{\mathrm{V}} \cos^2 \theta_s + E_{\mathrm{H}} \sin^2 \theta_s, \tag{8}$$

$$E_{\mathrm{QH}} = E_{\mathrm{V}} \sin^2 \theta_s + E_{\mathrm{H}} \cos^2 \theta_s, \tag{9}$$

where $E$ denotes a quantity that can be treated linearly in the radiative transfer; in this study, surface emissivity. Here, $E_{\mathrm{QV}}$ and $E_{\mathrm{QH}}$ represent the quasi-vertical and quasi-horizontal emissivities effectively observed by the instrument, while $E_{\mathrm{V}}$ and $E_{\mathrm{H}}$ correspond to the Earth-relative vertical and horizontal surface emissivities.

The scan angle $\theta_s$ is defined as the rotation angle of the scan mirror and is computed for each observation as

$$\theta_s = -z_{\text{swath}} + (\text{scanpos} - 1)\, z_{\text{FOV}}, \tag{10}$$

where $z_{\text{swath}}$ denotes the angular half-width of the swath and $z_{\text{FOV}}$ is the angular spacing between adjacent fields of view.

To account for this effect, a dedicated function, `compute_polarization_coeffs()`, was implemented in the sea-ice model. This function computes polarization-mixing coefficients, `pol0` and `pol1`, which weight the vertically and horizontally polarized emissivities, respectively. For conical scanners, the polarization geometry is fixed and the coefficients are set to $(\texttt{pol0}, \texttt{pol1}) = (1, 0)$ for vertically polarized channels and $(0, 1)$ for horizontally polarized channels. For cross-track scanners, the coefficients are computed using the $\cos^2\theta_s$ and $\sin^2\theta_s$ factors derived from the scan angle, ensuring physically consistent polarization mixing.

These coefficients are computed during dataset construction and passed to the `SeaiceEmisNN` layer, where they are used to combine the network-predicted $E_{\text{V}}$ and $E_{\text{H}}$. Importantly, this polarization mixing is prescribed from known instrument physics and is not learned by the neural network.

In the current implementation, the dependence of emissivity on zenith angle is neglected, implicitly assuming that frozen surfaces behave approximately as Lambertian emitters. While this approximation is reasonable as a first step, future work could include the zenith angle as an additional predictor in the `SeaiceEmisNN` layer to assess its impact on the simulated radiances and retrieval performance.

As previously demonstrated for AMSR2 (Geer, 2024a), the empirical sea-ice surface emissivity model trained here could also be employed within an atmospheric four-dimensional variational (4D-Var) data assimilation system. In this context, it would enable the assimilation of additional surface-sensitive microwave observations from cross-track sensors through the introduction of observation-space auxiliary control variables, or "sink variables", representing sea-ice concentration and the empirical sea-ice state parameters.

## 2.3 Building the dataset

Tables 2 and 3 provide an overview of the data used to train and run the hybrid ML model, all of which were retrieved from the ECMWF Meteorological Archival and Retrieval System (MARS) archive. In particular, Table 2 lists the original variables and provides their definitions and descriptions; in most cases, these variables are used directly as inputs to the hybrid model (see Figures 1 and 2). A smaller subset of the hybrid model inputs is derived from these variables and is summarized in Table 3.

The data used in this study can be grouped into three main categories: (i) satellite observations, (ii) output from the ECMWF IFS 12h background forecast, and (iii) output from the ECMWF IFS 6h analysis.

Variables in the first two categories were retrieved in `.odb` format, while those in the third

category were retrieved in `.grib` format. The `.odb` format is the native format of the ECMWF Observational DataBase (ODB), a relational database used to efficiently store, access, and process large volumes of meteorological observations and quantities generated during the DA process within the ECMWF IFS system.

Table 2: Summary of data used in the hybrid machine learning framework.

| **Variable** | **Description** | **Source** |
|---|---|---|
| obs. value | Brightness temperature measured by AMSU-A onboard METOP-B, METOP-C, NOAA-15, NOAA-18, and NOAA-19. Channels: 23.8, 31.4, 50.3, 52.8, and 89±1 GHz QV. | .odb (AMSU-A observations) |
| scan position | Position number along the scan. AMSU-A has 30 cross-track scan positions. | .odb (AMSU-A observations) |
| lat, lon | Latitude and longitude of the satellite observation (degrees). | .odb (AMSU-A observations) |
| date, time | Date and time of the satellite observation (YYYY-MM-DD HH:MM:SS). Used to construct `istep` (see Table 3 for details) | .odb (AMSU-A observations) |
| lsm | Land–sea mask from the ECMWF IFS, with values 0 for ocean and 1 for land. | .odb (ECMWF IFS 12h background) |
| tsfc | Surface (skin) temperature. | .odb (ECMWF IFS 12h background) |
| windspeed10m | 10 m wind speed. | .odb (ECMWF IFS 12h background) |
| cloud_fraction | Effective cloud fraction used in the atmospheric radiative transfer model. | .odb (ECMWF IFS 12h background) |
| emis_water | Ocean surface emissivity at AMSU-A channels, simulated using the SURFEM-Ocean model. | .odb (ECMWF IFS 12h background) |
| fg_dep | ECMWF IFS first-guess departure to observations. | .odb (ECMWF IFS 12h background) |
| tausfc_cld | Surface-to-space transmittance in the cloudy atmospheric column for each channel. | .odb (ECMWF IFS 12h background) |
| tausfc | Surface-to-space transmittance in the clear atmospheric column for each channel. | .odb (ECMWF IFS 12h background) |
| tup_cld | Upwelling atmospheric brightness temperature in the cloudy column for each channel. | .odb (ECMWF IFS 12h background) |

| | | |
|---|---|---|
| tup | Upwelling atmospheric brightness temperature in the clear atmospheric column for each channel. | .odb (ECMWF IFS 12h background) |
| tdown_cld | Downwelling atmospheric brightness temperature in the cloudy atmospheric column for each channel. | .odb (ECMWF IFS 12h background) |
| tdown | Downwelling atmospheric brightness temperature in the clear atmospheric column for each channel. | .odb (ECMWF IFS 12h background) |
| grid | Fixed positions from the N80 reduced Gaussian grid. | .grb (ECMWF IFS 6h analysis) |
| sic_initials | Sea-ice concentration from the NEMO-LIM3 model on the N80 reduced Gaussian grid. The 6-hourly data are monthly averaged and repeated daily. One additional day prior to training is included for time-lagging. | .grb (ECMWF IFS 6h analysis) |
| tsfc_initials | Surface (skin) temperature from the ECMWF IFS 6-hourly analysis on the N80 reduced Gaussian grid. Daily averages are used as initial conditions to constrain the sea-ice likelihood term. | .grb (ECMWF IFS 6h analysis) |

Table 3: Inputs in the hybrid machine learning framework derived from variables listed in Table 2.

| **Variable** | **Description** | **Derived from** |
|---|---|---|
| tsfc_norm | Normalized surface temperature, computed from the surface skin temperature as `tsfc_norm` $= \max(273.0 - \texttt{tsfc}, 0.0)/30.0$. The factor 30.0 is for normalization and is chosen so that maximum values do not much exceed 1, as commonly done in machine-learning applications. | tsfc |
| igrid | Subset of the N80 reduced Gaussian grid points defining the final fixed grid. This subset is constructed by mapping satellite observations to `grid` using a BallTree with a haversine metric and retaining only ocean grid points (based on `lsm`) with at least one associated observation. | grid, lsm, lat, lon |
| istep | Temporal index indicating the position of each observation within the training period. It is computed from the Julian day by discretizing time into daily steps (one step per day), with a temporal shift applied to align the index with the effective centre of the ECMWF 12h assimilation windows. | date, time |

| | | |
|---|---|---|
| sensor_id | Integer index identifying the satellite platform associated with each observation. A unique identifier is assigned to each sensor and repeated for all observations from that platform | Observation metadata |
| iobs | Unique identifier for each individual satellite observation. | Observation index |

- **Satellite observations**

  The satellite observations are provided by AMSU-A, a cross-track scanning microwave radiometer onboard multiple polar-orbiting platforms, including METOP-B, METOP-C, National Oceanic and Atmospheric Administration (NOAA) satellite-15 (retired on 19 August 2025), NOAA-18 (retired on 6 June 2025), and NOAA-19 (retired on 13 August 2025). AMSU-A samples 30 scan positions across the swath, with scan angles ranging from $-48.33°$ to $+48.33°$, where $0°$ corresponds to the nadir view, and an angular spacing between adjacent fields of view of $3.333°$.

  In this study, we focus on AMSU-A channels that are most sensitive to surface properties, namely 23.8 GHz QV, 31.4 GHz QV, 50.3 GHz QV, 52.8 GHz QV, and 89±1 GHz QV. Observations were initially retrieved for all available platforms over the period from 13 March 2024 to 6 June 2025, covering latitudes north of $50.5°$ N and south of $50.5°$ S. The experiments presented here, however, use only AMSU-A observations from METOP-B for the period from 1 April 2024 to 31 March 2025 and are restricted to the Northern Hemisphere. In addition, due to data quality issues, the 89 GHz QV channel from METOP-B has been excluded, and the final dataset therefore consists of the four lowest-frequency channels.

- **Output from the ECMWF IFS 12h background forecast**

  All atmospheric radiative transfer variables and some surface variables, collocated in space and time with each satellite observation, are retrieved from the ECMWF IFS 12h background forecast. For the period from 13 March 2024 to 12 November 2024, the ECMWF IFS Cycle 49r1 pre-operational configuration was used, while from 13 November 2024 onwards the operational ECMWF IFS Cycle 49r1 was employed.

- **Output from the ECMWF IFS 6h analysis**

  The fixed spatial grid and initial conditions for skin temperature and sea-ice concentration are provided by the ECMWF IFS 6h analysis. In particular, the reduced Gaussian N80 grid is adopted as the common reference grid for mapping satellite observations.

Figure 3 illustrates the workflow used to construct the dataset for training the hybrid ML model.

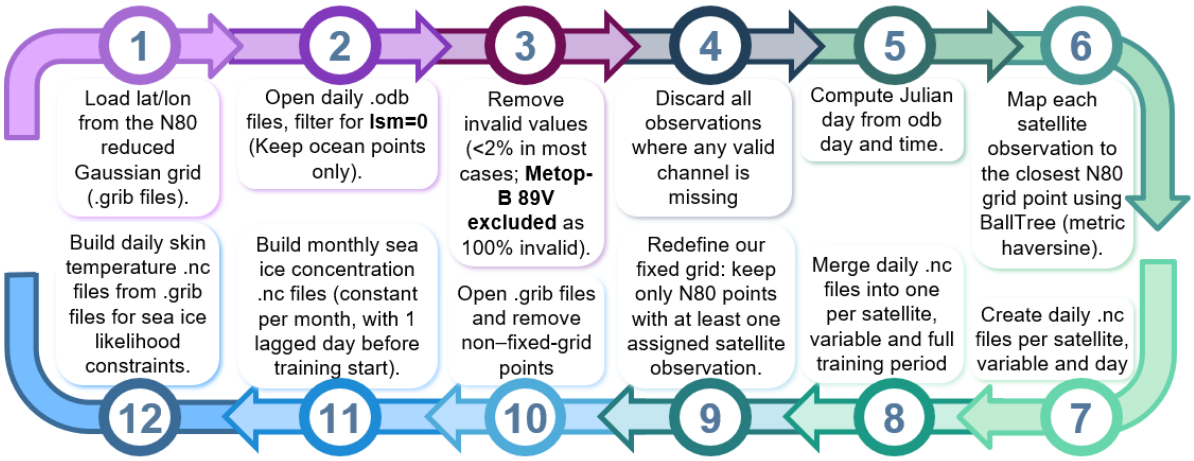


Figure 3: Workflow illustrating the steps followed to build the dataset for the hybrid machine learning model.

## 2.4 Running the model

Once the dataset was prepared, five experiments were conducted. Table 5 summarises the AMSU-A channels and the model parameters that are common to all experiments. Table 6 lists the experiment-specific model parameters, highlighting the values that vary between experiments.

Table 5: AMSU-A channels and model parameters common to all experiments.

| **Channel index** $j$ | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| Frequency (GHz) | 23.8 | 31.4 | 50.3 | 52.8 |
| Nominal polarization | QV | QV | QV | QV |
| Short name | 23V | 31V | 50V | 53V |
| **Model parameters common to all five experiments** | | | | |
| $c_{\mathrm{bgerr}}$ | 0.002 | | | |
| $\sigma_{\mathrm{bg}}$ | $1 \times 10^{-5}$ | | | |
| $C_{\mathrm{min}}$ | 0 | | | |
| $C_{\mathrm{max}}$ | 0.3 | | | |
| $\sigma_{\mathrm{false}}$ | 0.02 | | | |
| $s_{\mathrm{peak}}$ | 0.2 | | | |
| Number of epochs | 8 | | | |
| Batch size | 1024 | | | |
| Neural network layers | 2 | | | |
| Width of first layer | 7 | | | |

During the Experiment 1, which serves as a control experiment, the hybrid ML model was applied without the angle-dependency extension, effectively treating AMSU-A as a conical-scanning sensor rather than a cross-track instrument. All model parameters were set to values similar to those used in Geer (2024a).

Table 6: Experiment-specific model parameters.

| **Experiment** | **Parameter** | 23V | 31V | 50V | 53V |
|---|---|---|---|---|---|
| Experiment 1 | $r$ (obs. error) | 2.5 | 3.5 | 8.5 | 8.0 |
| | $b_{\text{ice,bg}}$ | 2.5 | 2.5 | 2.5 | 2.5 |
| | $b_{\text{wat,bg}}$ | 5.0 | 5.0 | 5.0 | 5.0 |
| | $b_{\text{bgerr}} = 0.001,\ e_{\text{bg}} = 0.8,\ \sigma_{\text{false}} = 0.02 \rightarrow 0.002$ (from epoch 3) | | | | |
| Experiment 2 | $r$ (obs. error) | 4.5 | 5.0 | 5.0 | 2.0 |
| | $b_{\text{ice,bg}}$ | 2.5 | 2.5 | 2.5 | 2.5 |
| | $b_{\text{wat,bg}}$ | 5.0 | 5.0 | 5.0 | 5.0 |
| | $b_{\text{bgerr}} = 0.001,\ e_{\text{bg}} = 0.8,\ \sigma_{\text{false}} = 0.02 \rightarrow 0.002$ (from epoch 3) | | | | |
| Experiment 3 | $r$ (obs. error) | 4.5 | 5.0 | 5.0 | 2.0 |
| | $b_{\text{ice,bg}}$ | 0.0 | 0.0 | 0.0 | 0.0 |
| | $b_{\text{wat,bg}}$ | 0.0 | 0.0 | 0.0 | 0.0 |
| | $b_{\text{bgerr}} = 1.0,\ e_{\text{bg}} = 0.7,\ \sigma_{\text{false}} = 0.02$ | | | | |
| Experiment 4 | $r$ (obs. error) | 4.5 | 5.0 | 5.0 | 2.0 |
| | $b_{\text{ice,bg}}$ | 0.0 | 0.0 | 0.0 | 0.0 |
| | $b_{\text{wat,bg}}$ | 0.0 | 0.0 | 0.0 | 0.0 |
| | $b_{\text{bgerr}} = 1.0,\ e_{\text{bg}} = 0.7,\ \sigma_{\text{false}} = 0.02$; angle dependency enabled | | | | |
| Experiment 5 | $r$ (obs. error) | 4.5 | 5.0 | 5.0 | 2.0 |
| | $b_{\text{ice,bg}}$ | 0.0 | 0.0 | 0.0 | 0.0 |
| | $b_{\text{wat,bg}}$ | 0.0 | 0.0 | 0.0 | 0.0 |
| | $b_{\text{bgerr}} = 1.0,\ e_{\text{bg}} = 0.6,\ \sigma_{\text{false}} = 0.02$; angle dependency enabled | | | | |

In the Experiment 2, the same configuration was retained, except for the observation errors $r$ assigned to the different channels. These were adjusted based on the standard deviation of AMSU-A BT differences between observations and simulations over the period from 1 April 2024 to 31 March 2025, as obtained from the Experiment 1.

In the Experiment 3, the revised observation error values from the Experiment 2 were retained, while several other model parameters were updated. In the first two experiments, following Geer (2024a), the penalty against false sea-ice concentration was strengthened from the third epoch onward by reducing $\sigma_{\text{false}}$ from 0.02 to 0.002, thereby enforcing a stronger physical constraint during the later stages of training. For this experiment and all subsequent ones, $\sigma_{\text{false}}$ was kept constant at 0.02 throughout all epochs. Also, in Geer (2024a), the background bias terms $b_{\text{ice,bg}}$ and $b_{\text{wat,bg}}$ were set to 2.5 K and 5.0 K, respectively, for all channels, reflecting the well-known scene-temperature–dependent biases of AMSR2. Such prior information is appropriate for AMSR2, for which these biases are well characterised and routinely accounted for in Numerical Weather Prediction (NWP) through variational bias correction schemes. For AMSU-A, however, no comparable a priori bias values are available. Although bias corrections are applied to AMSU-A radiances in NWP applications, they are estimated adaptively within the DA system, and their validity—particularly over sea-ice surfaces—remains unclear. For these reasons, in this experiment and all subsequent ones, zero prior biases were adopted for AMSU-A, allowing the training process to estimate the bias corrections directly. To prevent unrealistic compensation for large BT contrasts between ocean and sea-ice, the magnitude of the bias term was constrained using a relatively small background error of $b_{\text{bgerr}} = 1$ K. The formulation of the emissivity regularization term $J_{\text{emis}}$ (Equation 7) also differs from that in Geer (2024a). In that study, the sea-ice emissivity model consisted of a single dense linear neural network layer, and the constraint was applied to the bias term of the lowest-frequency channel emissivity (10.65 GHz). In the present work, emissivity is predicted using two dense neural network layers, and the constraint is instead imposed directly on the predicted emissivity of the lowest-frequency channel (23.8 GHz). Accordingly, the parameter $e_{\text{bg}}$ defines a lower bound on the emissivity at this channel. For this experiment, $e_{\text{bg}}$ was set to 0.7, slightly lower than the 0.8 used in previous experiments, while the same background error $\sigma_{\text{bg}}$ was retained to control the sharpness of the penalty when emissivities fall below this threshold.

The Experiment 4 uses exactly the same parameter settings as the Experiment 3, but with the angle-dependency extension enabled for the first time. This activates the cross-track formulation of the model described in Section 2.2.

The Experiment 5 retains the same configuration as the fourth experiment, with the only modification being a further reduction of the emissivity lower bound to $e_{\text{bg}} = 0.6$.

Table 7 summarizes the network architecture used in all these experiments. The training period spans from 1 April 2024 to 31 March 2025. The fixed grid used to map the satellite observations is based on the N80 reduced Gaussian grid, restricted to latitudes north of 50.5°N and excluding land points using the ECMWF IFS land–sea mask, resulting in a total of 2482 grid points.

Among the trainable parameters of the hybrid ML model are the empirical sea-ice properties. As in Geer (2024a), three such properties are used; in that study, all three were

Table 7: Summary of dataset characteristics, trainable parameters, and fixed parameters used in this study.

| | |
|---|---|
| **Training data** | |
| Time period start | 1 April 2024 |
| Time period end | 31 March 2025 |
| Number of observations | 2,533,103 |
| Channels per observation | 4 |
| **Total observations by channels** | **10,132,412** |
| **Geographical grid** | |
| Number of spatial grid points | 2482 |
| Number of time points | 365 or 366 |
| **Trainable parameters** | |
| sea-ice concentration map $C_{\text{ice}}$ (366 days) | 908,412 |
| Empirical micro- and macro-physical properties observation space $X_{\text{ice}_{\text{obs}}}$ (1 variables, nb obs) | 2,533,103 |
| Empirical micro- and macro-physical properties grid space $X_{\text{ice}_{\text{grid}}}$ (2 variables, 365 days) | 1,811,860 |
| sea-ice emissivity model weights $w$ | 58 |
| Observation bias correction coefficients ($b_{\text{ice}}$, $b_{\text{wat}}$) | 8 |
| Ocean surface emissivity wind-speed correction coefficients $b_{\text{emis}}$ | 4 |
| **Total trainable parameters** | **5,253,445** |
| **Fixed parameters (per observation)** | |
| Atmospheric radiative transfer variables | 25 |
| Surface variables (tsfc, $tsfc_{norm}$, windspeed10m, emis_water) | 7 |
| Geolocation variables (grid point, day, observation number) | 3 |

represented as daily maps on the fixed grid (grid space). In the present implementation, two properties remain on the fixed grid, while the third varies on a per-observation basis (observation space) during training, providing additional flexibility to address limitations such as the inability to represent diurnal variability in optical properties over summer sea-ice, potentially associated with freeze–thaw cycles.

Sea-ice concentration is also treated as a freely adjustable variable and estimated as daily maps on the fixed grid, spanning 366 days to incorporate information from both the current and previous day at each observation location. Weights of 0.6 and 0.4 were applied to the current and previous day, respectively, giving greater importance to the current day's ice concentration. This approach provides a simple approximation to a persistence sea-ice forecast model, constraining the temporal evolution of the geophysical state. Such constraints are a key component of DA systems, and while a fully physical sea-ice model would be preferable, persistence-based approaches have proven effective in practice (Buehner et al., 2013). In particular, sea-ice persistence helps reduce spurious ice features introduced by cloud contamination over open ocean, under the assumption that sea-ice evolves more slowly than atmospheric cloud structures.

Sea-ice emissivity is modelled using a fully connected neural network consisting of two dense layers. In this work, we choose an architecture with 7 neurons in the first layer and 2 neurons in the second layer. Given the four channels used, this design results in a total of 58 trainable parameters in the emissivity network. Wind-speed-dependent emissivity corrections and observation bias corrections are also treated as channel dependent, with the latter including separate bias terms over ice and water. This adds a further 4 and 8 trainable bias parameters, respectively.

Training of the hybrid model and the associated geographical fields was performed on the ECMWF Atos supercomputer and no Graphics Processing Unit (GPU)s were required. The model was trained for eight epochs, which was sufficient for the present application and helped to limit overfitting. The cost function (Equation 1 ) was minimized using the Adam optimizer (Kingma and Ba, 2015) with default settings, including a learning rate of 0.001, and a mini-batch size of 1024.

The Python code supports both training from scratch and inference using a pre-trained model. In inference mode, the model parameters obtained from a previous training run are kept fixed, and the model is applied in forward mode to estimate sea-ice concentration and the other model variables. During inference, the dataset is split into three subsets due to numerical stability issues encountered when applying `model.predict()` to the full dataset at once. Predictions are therefore computed independently for each subset and subsequently concatenated. From the perspective of the Keras framework, fixed model parameters are treated as input features, while the observations are treated as output targets.

The performance of the hybrid ML model is evaluated over the one-year training period using the mean, standard deviation, and Root Mean Square Error (RMSE) of the differences between observed and simulated BTs as a function of scan position. Statistics are computed for all observations, for near-nadir views (scan positions 14, 15, and 16), and for large scan angles (scan positions 0–5 and 25–30). These diagnostics are compared with the corresponding FG departures from the ECMWF IFS to assess the added value of the

hybrid ML model relative to the current operational approach. For spatial illustration, most figures show results for a single arbitrarily chosen day from the training period (1 April 2024), corresponding to the first day of training.

# 3 Results

The results presented here primarily focus on analysing the differences between all conducted experiments in terms of BTs, with the aim of identifying the model configuration that minimises discrepancies between observed and simulated BTs (Sections 3.1 and 3.2). In addition, one of the key objectives of this study was to extend the hybrid ML framework to handle cross-track scanning sensors; for this reason, Section 3.3 is fully dedicated to this aspect. Finally, Section 3.4 primarily serves to showcase variables obtained during the training of the hybrid model beyond BTs. These variables provide valuable insights for refining the model and identifying areas for future improvement, although only preliminary analyses are presented here. This includes an initial examination of the loss terms during training for Experiments 1, 4, and 5, as well as maps of sea-ice concentration, empirical sea-ice properties, and sea-ice emissivity obtained for Experiment 4. Additionally, the sea-ice emissivity is compared for two different locations in the Arctic.

## 3.1 Control experiment

During the EUMETSAT conference in Lyon (15–19 September 2025), where I was honoured with an Early Career Scientist Award, I presented the first preliminary results (González-Flórez et al., 2025). At that time, it was the first occasion I was able to fully run the code and train the model. Upon returning from the conference, I realized that the code was not fully deterministic, and each run produced slightly different results. To address this, I added a fixed random seed to ensure reproducibility. Consequently, the results shown here for the Experiment 1 differ slightly from those presented at the conference, although the same model parameters, settings, and dataset were used, and the conclusions remain consistent. Table 8 presents the mean bias, standard deviation, and RMSE by scan position over the full one-year training period for the Experiment 1, which serves as the control experiment throughout this study. The differences between observed and simulated BTs are smaller for near-nadir positions than for large viewing angles, suggesting that explicitly accounting for viewing-angle dependency may improve model performance.

Figure 4 is an updated version of Figure 4 from the poster presented at the EUMETSAT conference (González-Flórez et al., 2025), based on the same 1 April 2024 data. It includes two additional rows: the third row shows the observed minus simulated BT differences, and the fourth row displays the ECMWF IFS FG departures to observations. The first two rows and the last two rows of the figure share colorbars to facilitate consistent visual interpretation. Simulated BTs reproduce the spatial patterns and absolute values of the observations well, particularly for the first two channels (panels a, b, e, and f), even when all scan positions are included. The third channel captures recognizable spatial patterns but with larger deviations (panels c and g), while the fourth channel exhibits the largest differences (panels d and h).

These day-specific results are supported by the annual statistics presented in Table 8.

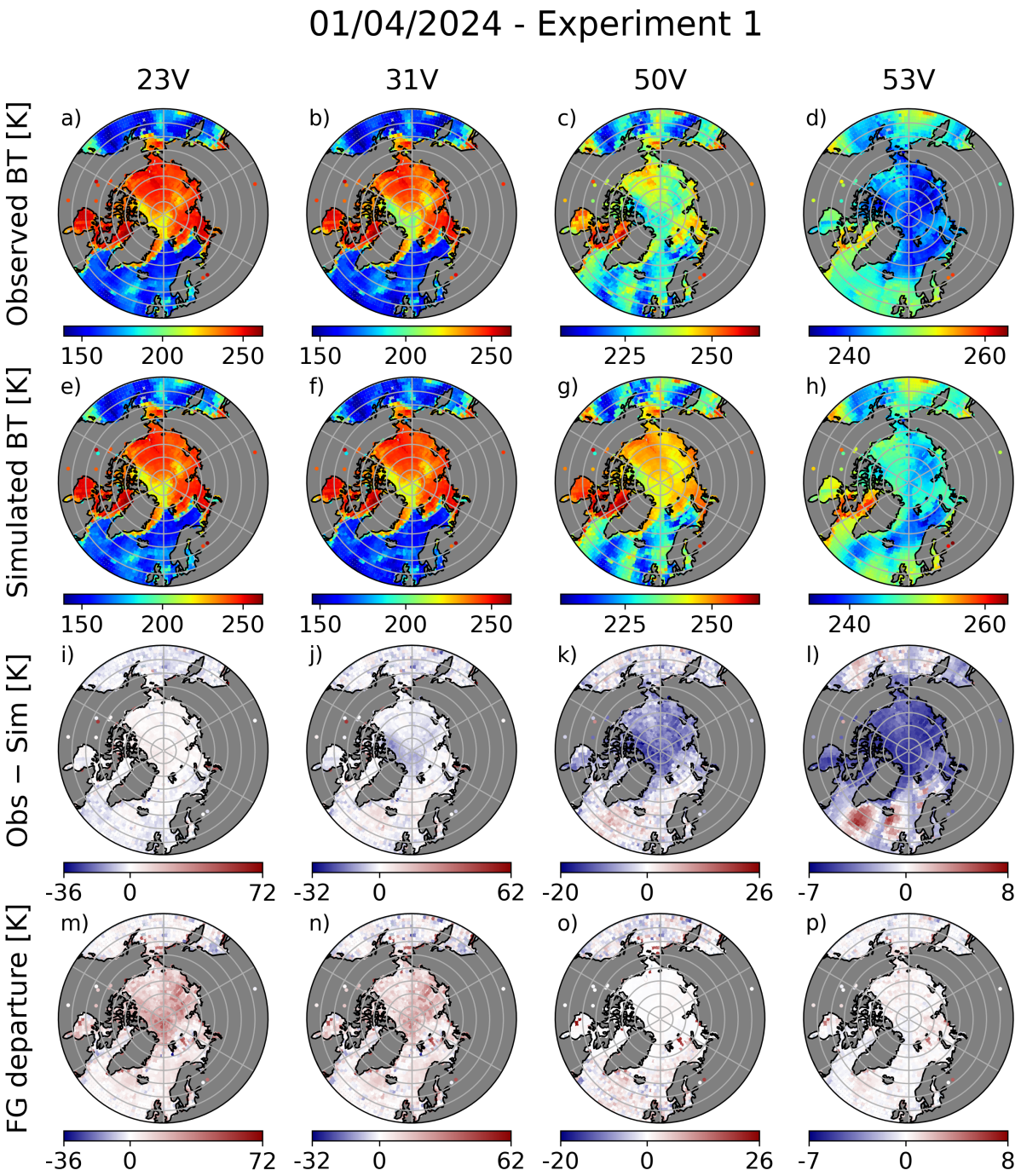


Figure 4: AMSU-A brightness temperatures (in K) over the Arctic on 1 April 2024 for the first (control) experiment. Columns correspond to the 23V, 31V, 50V, and 53V channels. Panels (a–d) show the observed brightness temperatures, (e–h) the simulated brightness temperatures, (i–l) the observed minus simulated (Obs−Sim) differences, and (m–p) the ECMWF IFS first-guess departures to observations. The first two rows share a common colour scale, as do the last two rows.

Table 8: Observed minus simulated brightness temperature (Obs−Sim) statistics for the first (control) experiment, computed over the full one-year period. Mean bias, standard deviation, and RMSE (in K) are reported for all scan positions, near-nadir views (14–16), and large scan angles (0–5 and 25–30). Values in brackets [†] denote the corresponding ECMWF IFS FG departures to observations for comparison.

| Statistic | Scan positions | 23V | 31V | 50V | 53V |
|---|---|---|---|---|---|
| (Obs−Sim).mean() (K) | All | 0.12 [9.83†] | -1.59 [6.84] | -4.09 [1.00] | -2.74 [0.14] |
| | Around nadir | -0.09 [6.82] | -1.31 [4.68] | -2.35 [1.33] | -1.41 [0.36] |
| | Large angles | 0.32 [12.11] | -1.74 [8.56] | -5.25 [0.81] | -3.57 [-0.03] |
| (Obs−Sim).std() (K) | All | 4.86 | 5.33 | 5.27 | 2.23 |
| | Around nadir | 4.52 | 4.81 | 5.04 | 2.64 |
| | Large angles | 5.13 | 5.78 | 5.25 | 1.67 |
| RMSE (K) | All | 4.86 | 5.56 | 6.66 | 3.53 |
| | Around nadir | 4.52 | 4.98 | 5.56 | 2.99 |
| | Large angles | 5.14 | 6.03 | 7.42 | 3.94 |

For the two lowest-frequency channels (23V and 31V), which are most sensitive to surface conditions, the mean Obs−Sim differences are substantially reduced relative to the ECMWF IFS FG departures to observations (values in brackets in the table). Considering all scan positions, the reductions correspond to approximately 99% for 23V and 77% for 31V. Notably, this improvement was achieved using data from a different satellite instrument (AMSU-A) than the one for which the hybrid ML model was originally developed (AMSR2), without extensive parameter tuning or major modifications to the model configuration, demonstrating the model's generality. For this specific day, the largest differences between simulated and observed BTs occur in coastal regions at the interface between land and sea, as well as over lakes (Figure 4, panels i and j). These areas warrant further investigation to identify the causes of the discrepancies and to explore potential improvements.

## 3.2 Sensitivity tests on model parameters

Table 9 is analogous to Table 8, but summarizes results for Experiments 2–5. Figure 5 presents observed and simulated BTs for all experiments on 1 April 2024, using common colour scales per channel, while Figure 6 additionally shows the observed minus simulated differences together with the ECMWF IFS FG departures to observations.

As described in Section 2.4, Experiments 1 and 2 differ only in the values of the observation errors. A comparison of Table 8 with Table 9 (Experiment 2) shows that this change alone reduces the mean Obs−Sim differences by at least 0.5 K for all channels except the first. However, when focusing on the single-day case of 1 April 2024, the third and fourth channels still exhibit relatively large absolute deviations from the observations (Figure 5, panels g, h, k and l vs. panels c and d) and are not yet able to fully reproduce the observed spatial patterns.

In Experiment 3, the observation error values from Experiment 2 were retained and, as justified in Section 2.4, the parameters $b_{\mathrm{ice,bg}}$, $b_{\mathrm{wat,bg}}$, $b_{\mathrm{bgerr}}$, and $e_{\mathrm{bg}}$ were updated. These adjustments led to a substantial reduction in the differences between observed and

Table 9: Observed minus simulated brightness temperature (Obs−Sim) statistics for the Experiments 2 to 5, computed over the full one-year period. Mean bias, standard deviation, and RMSE (in K) are shown for all scan positions, near-nadir positions (14–16), and large-angle positions (0–5 and 25–30). Values in brackets [†] indicate ECMWF IFS first-guess departures to observations for comparison.

| Experiment | Statistic | Scan position | 23V | 31V | 50V | 53V |
|---|---|---|---|---|---|---|
| Experiment 2 | (Obs−Sim). mean() (K) | All | 1.27 [9.83†] | -0.61 [6.84] | -3.54 [1.00] | -2.23 [0.14] |
| | | Nadir | 1.02 [6.82] | -0.39 [4.68] | -1.76 [1.33] | -1.00 [0.36] |
| | | Large angles | 1.46 [12.11] | -0.75 [8.56] | -4.75 [0.81] | -3.00 [-0.03] |
| | (Obs−Sim). std() (K) | All | 5.86 | 5.47 | 4.82 | 1.81 |
| | | Nadir | 5.44 | 4.99 | 4.50 | 2.05 |
| | | Large angles | 6.15 | 5.90 | 4.86 | 1.36 |
| | RMSE (K) | All | 5.99 | 5.51 | 5.98 | 2.87 |
| | | Nadir | 5.54 | 5.01 | 4.83 | 2.28 |
| | | Large angles | 6.32 | 5.95 | 6.79 | 3.30 |
| Experiment 3 | (Obs−Sim). mean() (K) | All | -0.39 | -0.14 | -0.22 | -0.03 |
| | | Nadir | -0.69 | -0.15 | 0.82 | 0.33 |
| | | Large angles | -0.13 | -0.09 | -0.92 | -0.29 |
| | (Obs−Sim). std() (K) | All | 5.26 | 5.32 | 3.92 | 0.80 |
| | | Nadir | 4.67 | 4.86 | 3.95 | 0.91 |
| | | Large angles | 5.72 | 5.72 | 3.84 | 0.68 |
| | RMSE (K) | All | 5.28 | 5.33 | 3.93 | 0.80 |
| | | Nadir | 4.72 | 4.86 | 4.04 | 0.97 |
| | | Large angles | 5.72 | 5.72 | 3.95 | 0.74 |
| Experiment 4 | (Obs−Sim). mean() (K) | All | -0.41 | -0.11 | -0.17 | -0.03 |
| | | Nadir | -0.80 | -0.31 | 0.72 | 0.31 |
| | | Large angles | -0.05 | 0.10 | -0.76 | -0.26 |
| | (Obs−Sim). std() (K) | All | 5.16 | 5.24 | 3.83 | 0.77 |
| | | Nadir | 4.62 | 4.73 | 3.88 | 0.88 |
| | | Large angles | 5.58 | 5.67 | 3.76 | 0.66 |
| | RMSE (K) | All | 5.17 | 5.25 | 3.83 | 0.77 |
| | | Nadir | 4.69 | 4.74 | 3.94 | 0.94 |
| | | Large angles | 5.58 | 5.67 | 3.84 | 0.71 |
| Experiment 5 | (Obs−Sim). mean() (K) | All | -0.55 | -0.16 | -0.12 | -0.03 |
| | | Nadir | -0.89 | -0.34 | 0.78 | 0.32 |
| | | Large angles | -0.22 | 0.06 | -0.70 | -0.27 |
| | (Obs−Sim). std() (K) | All | 5.16 | 5.25 | 3.72 | 0.75 |
| | | Nadir | 4.63 | 4.73 | 3.81 | 0.86 |
| | | Large angles | 5.56 | 5.68 | 3.63 | 0.64 |
| | RMSE (K) | All | 5.19 | 5.26 | 3.72 | 0.75 |
| | | Nadir | 4.72 | 4.74 | 3.89 | 0.92 |
| | | Large angles | 5.57 | 5.68 | 3.70 | 0.69 |

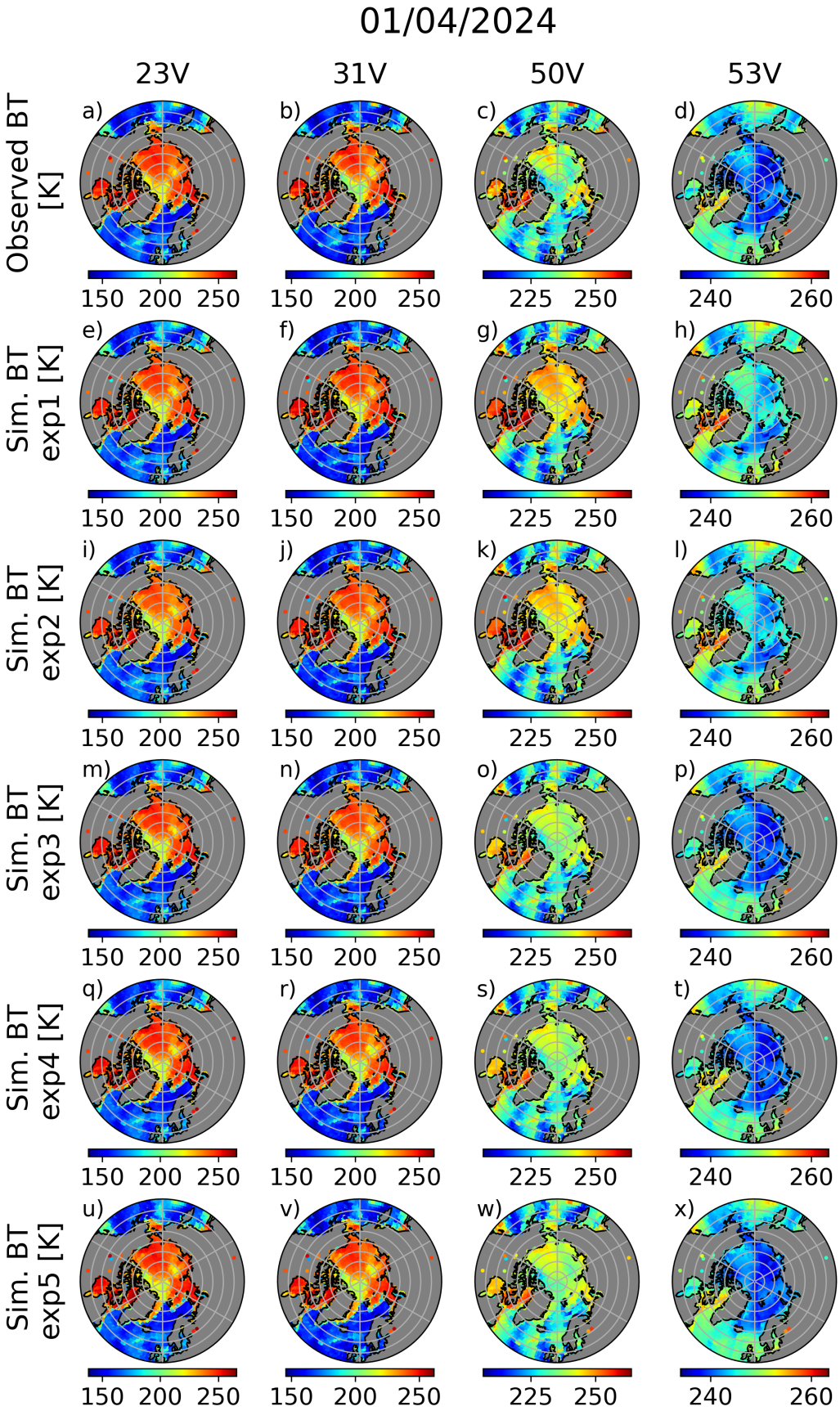


Figure 5: Observed and simulated AMSU-A brightness temperatures (in K) over the Arctic on 1 April 2024, shown for the 23V, 31V, 50V, and 53V channels. The top row presents observations, with subsequent rows showing simulations from each experiment using a common colour scale per channel.

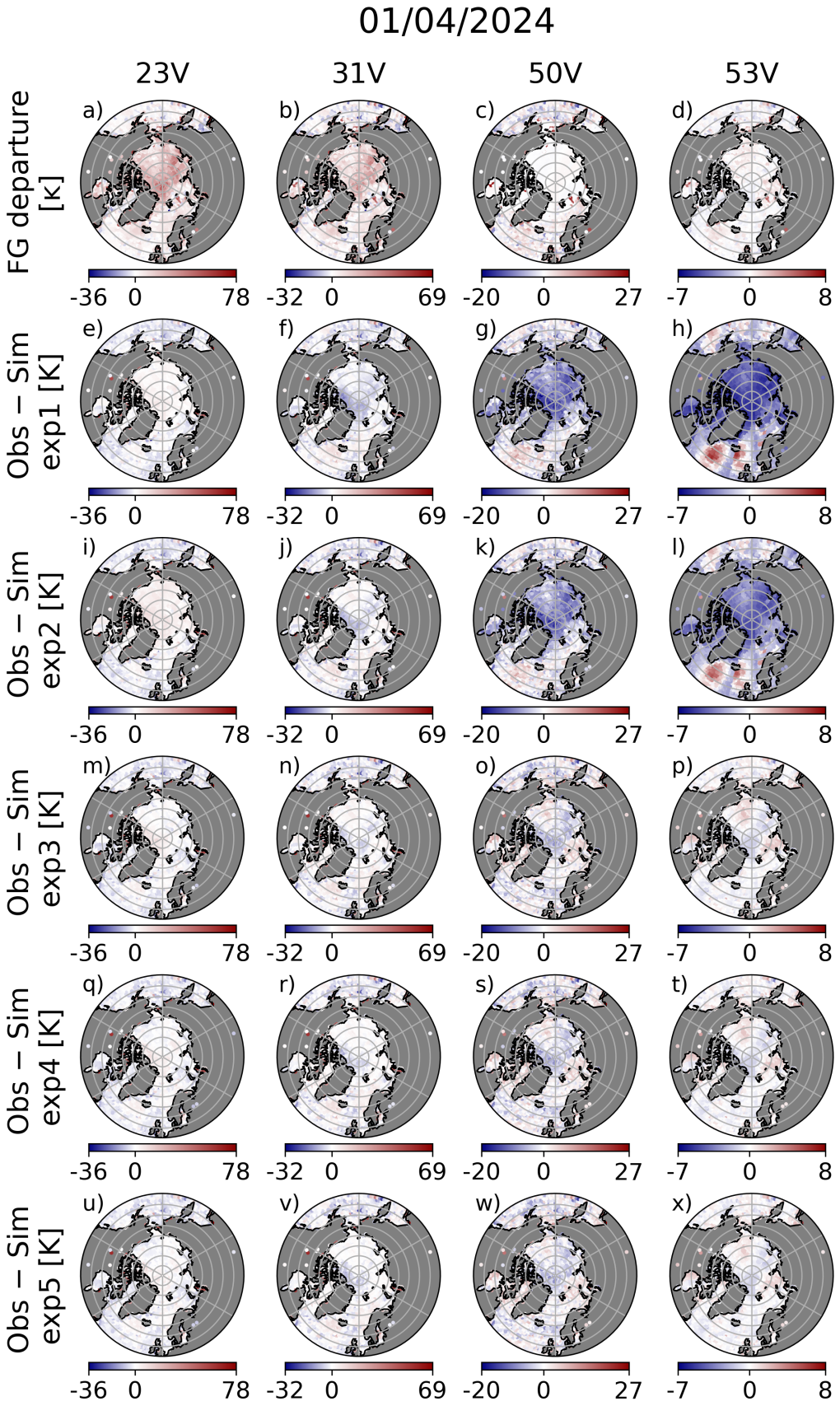


Figure 6: ECMWF IFS first-guess departures to observations (top row) and differences between observed and simulated AMSU-A brightness temperatures (in K) for each experiment (subsequent rows) over the Arctic on 1 April 2024, shown for the 23V, 31V, 50V, and 53V channels with a common colour scale per channel.

simulated BTs relative to the first two experiments (Table 8 and Table 9, Experiments 2 and 3). Considering all scan positions, the mean Obs−Sim differences for the third and fourth channels in Experiment 3 decrease by more than 3 K and 2 K, respectively, compared to Experiment 2. Moreover, comparing Experiments 1 and 3, this stepwise parameter tuning reduces mean Obs−Sim differences by roughly 91% for 31V, 95% for 50V, and 99% for 53V, while the 23V channel, already well reproduced in the Experiment 1, remains largely unchanged.

This improvement is also evident in Figure 5, where these channels now reproduce the observed spatial patterns more consistently. Accordingly, panels o and p of Figure 6 display a larger extent of near-zero differences (white or pale colours) than panels k and l.

Finally, Experiments 4 and 5 differ only in the value of the emissivity lower bound, which is reduced from $e_{\mathrm{bg}} = 0.7$ to $e_{\mathrm{bg}} = 0.6$. This change has a very limited impact on the results: the mean Obs−Sim difference remains unchanged for the fourth channel, decreases by 0.05 K for the third channel, and increases by 0.14 K and 0.05 K for the first and second channels, respectively (Table 9, Experiments 4 and 5). These differences are very small, and in Figures 5 and 6 the resulting patterns are almost indistinguishable.

A comparison between the simulated minus observed BT differences and the ECMWF IFS FG departures to observations in Tables 8 and 9 shows that, only in Experiments 1 and 2, the ECMWF IFS FG departures to observations are smaller than the mean Obs−Sim differences and only for the third and fourth channels. For Experiments 3, 4, and 5, the mean Obs−Sim differences are lower than the corresponding ECMWF IFS FG departures to observations across all channels, with reductions exceeding 6 K for the first two channels.

## 3.3 Angle dependency

The inclusion of viewing-angle dependency warrants a dedicated section, as it constitutes the principal methodological advancement implemented in the hybrid ML model during the first year of this fellowship. Experiments 3 and 4 provide a clean framework to assess its impact, since they differ only in the inclusion of angle dependency while all other model parameters are kept identical.

When considering only large viewing angles observations, the inclusion of angle dependency leads to a slight improvement in model performance for some channels (Table 9, Experiments 3 and 4). In particular, the mean Obs−Sim differences are reduced by approximately 60% for the first channel (from -0.13 to -0.05 K) and by about 17% (from -0.92 to -0.76 K) for the third channel. For the fourth channel, the mean differences remain very similar in both experiments. In addition, both the standard deviation and the RMSE exhibit reductions of approximately 1–4% across the four channels, indicating slightly more consistent model behaviour at large scan angles when including the angle dependency.

Interestingly, when the angle dependency is included and the analysis is restricted to near-nadir observations, the mean Obs−Sim differences increase slightly—by a few tenths of a kelvin—despite modest reductions in the standard deviation and RMSE. This behaviour warrants further investigation, as the angle-dependent formulation appears to enhance performance at larger viewing angles while introducing a small degradation under near-

nadir conditions.

Figure 7 shows the number of observations per scan position for a single channel (the distribution is identical for all channels). The left panel corresponds to the entire training period, while the right panel shows the distribution for 1 April 2024 only. Both panels exhibit the same overall behaviour: the number of observations increases towards the scan edges, with a pronounced maximum at the right edge (scan positions above 25), and decreases near nadir angles. This reflects the typical angular sampling of the instrument.

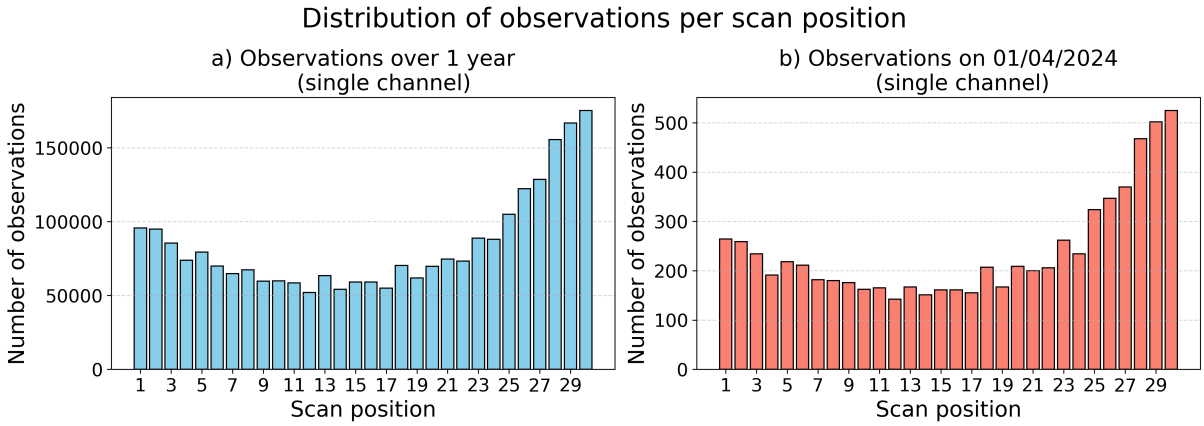


Figure 7: Distribution of AMSU-A observations onboard METOP-B per scan position for the 23.8 GHz (23V) channel over the Northern Hemisphere (latitudes north of 50.5° N). a) Histogram for the entire training period (1 April 2024 to 31 March 2025). b) Histogram for 1 April 2024.

Figure 8 further illustrates this behaviour by showing the geographical distribution of observed minus simulated BTs on 1 April 2024, separated by scan position and experiment. The first two rows show observations acquired close to nadir (scan positions 14, 15 and 16) for Experiments 3 and 4, respectively, while the last two rows display observations from the scan edges (scan positions 0–5 and 25–30) for the same experiments. Notably, far fewer observations are acquired near nadir, as individual satellite overpasses on that day can be clearly distinguished.

All panels for a given channel share the same colour scale, which reveals an interesting and consistent behaviour: for the third and fourth channels (Figure 8, panels c, d, g and h), the observed minus simulated BT differences for near-nadir observations tend to be predominantly positive (red colours), indicating that the simulation tends to underestimate the BTs (i.e., exhibits a cold bias). In contrast, for large scan angles, negative differences (blue colours) are more frequent (Figure 8, panels k, l, o, and p), suggesting a relative warm bias in the simulation at these angles. This pattern is also reflected in the statistics computed over the full one-year training period (Table 9, Experiments 3 and 4), supporting the robustness of the findings at both near-nadir and large scan angles.

On 1 April 2024, the largest positive observed minus simulated BT differences for near-nadir observations are found over the Novaya Zemlya archipelago and the western part of Hudson Bay (Figure 8, panels c, d, g and h). In contrast, the strongest negative differences for large scan angles are mainly located over the western Arctic Ocean (Figure 8, panels k, l, o and p), a region predominantly covered by multi-year sea-ice (Comiso, 2012).

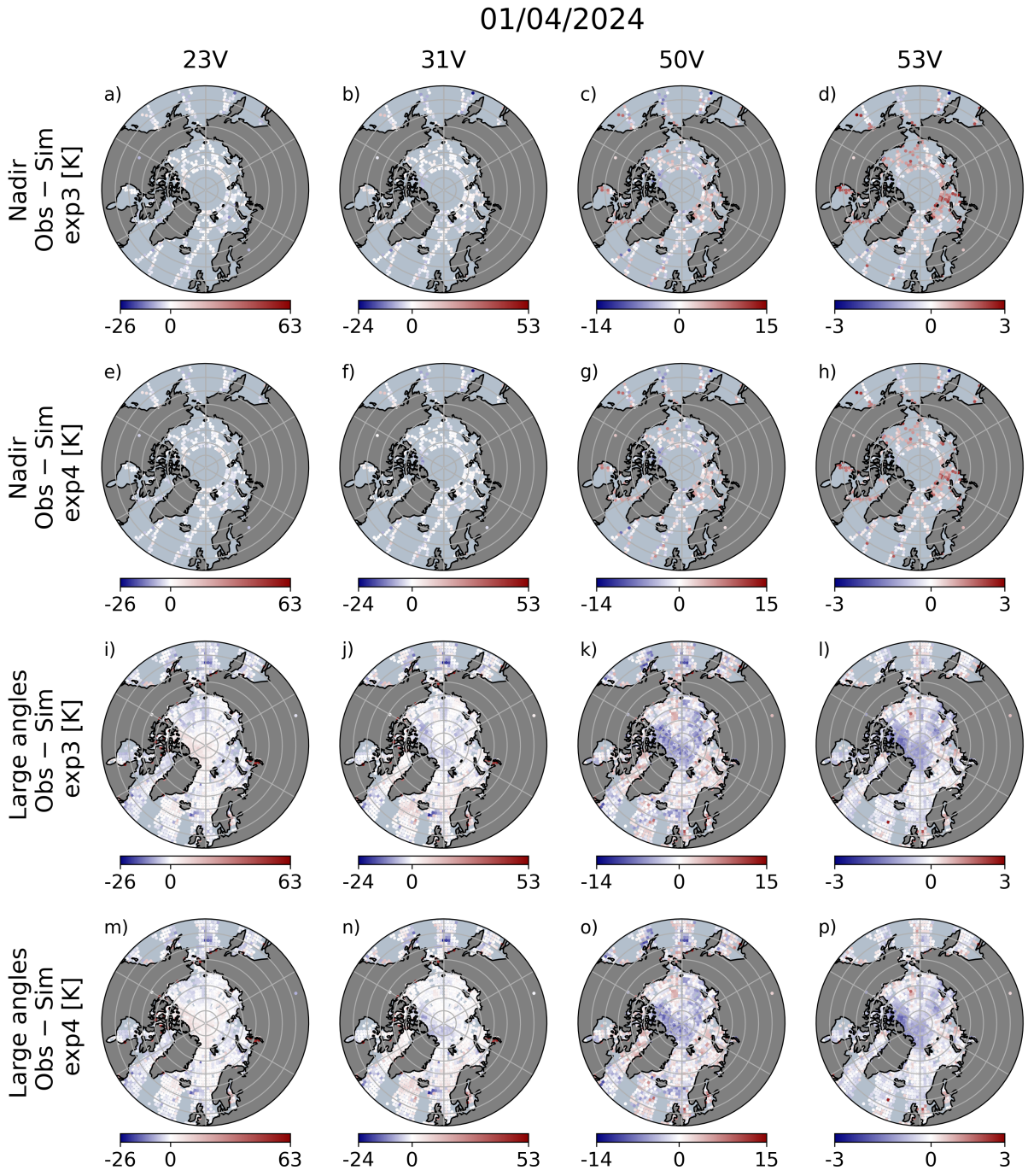


Figure 8: Differences between observed and simulated AMSU-A brightness temperatures (in K) from Experiment 3 (neglecting angle dependency) and Experiment 4 (incorporating angle dependency) over the Arctic on 1 April 2024. Near-nadir views (scan positions 14, 15 and 16) are shown in the first two rows, and large scan angles (scan positions 0–5 and 25–30) in the subsequent rows. Results are presented for the 23V, 31V, 50V, and 53V channels, using a common colour scale for each channel.

In general, the inclusion of physically based constraints, such as viewing-angle dependency, is expected to enhance model realism. Identifying the conditions and regions under which this additional complexity leads to improvements or degradations in performance will be important for guiding future model development and requires further investigation.

In summary, both for this specific day and when considering the annual statistics, the differences between observed minus simulated BTs obtained with and without angle dependency remain subtle. The transition from Experiment 2 to Experiment 3—which primarily involved tuning the observation errors together with $b_{\mathrm{ice,bg}}$, $b_{\mathrm{wat,bg}}$, $b_{\mathrm{bgerr}}$, and $e_{\mathrm{bg}}$—resulted in a larger improvement in absolute terms than that achieved by introducing the angle dependency alone, although it is too early to assess the overall value of including angle dependency, and further investigation is required.

## 3.4 Additional exploratory analyses

Figure 9 shows, as an example, the total loss $J$ (the objective function minimized during training) and its constituent terms as a function of the training epoch for Experiments 1, 4, and 5. As mentioned earlier, eight epochs were selected for these experiments. The loss terms span a wide range of magnitudes; therefore, all terms except the total and observation losses have been rescaled to make them visible in the figure. For all three experiments, the total loss decreases sharply during the first four epochs, after which the reduction becomes much more gradual.

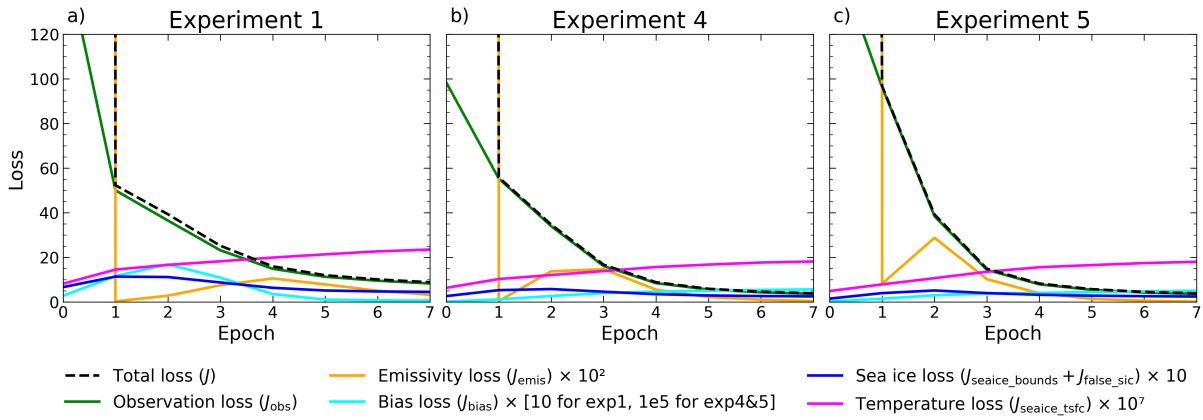


Figure 9: Evolution of the different loss components as a function of training epoch for (a) Experiment 1 , (b) Experiment 4, and (c) Experiment 5. Selected loss components are rescaled for visualization.

Experiment 1 (Figure 9a) is the closest to the experiment conducted by Geer (2024a), as we adopt the same or very similar values for all model parameters and do not include the angle dependency. It therefore serves as an initial reference configuration, ensuring that the model is correctly implemented and generates realistic results. The main differences lie in the use of AMSR2 data in Geer (2024a) versus AMSU-A data in this study, as well as in the time period considered. A comparison between Figure 7 in Geer (2024a) and Figure 9a shows that, in both cases, the total loss at epoch zero is off-scale. At epoch one, Geer (2024a) reports a total loss around 40, whereas in our experiments it is approximately 50 for Experiment 1, 55 for Experiment 4, and 95 for Experiment 5. By epoch four, Geer (2024a)'s total loss stabilizes around 15, while in our case Experiment 1

stabilizes around 10, and Experiment 4 and Experiment 5 even reach lower values.

Regarding the emissivity loss, $J_{emis}$ (Equation 7), which enforces a physical lower bound on the sea-ice emissivity for the lowest-frequency channel, the values in the present experiments are higher and the curve trends differ from those reported by Geer (2024a). Note that in Figure 9, $J_{emis}$ is scaled by $10^2$, whereas Geer (2024a) uses $10^3$. However, as mentioned in Section 2.4, the constraint here is applied directly to the predicted emissivity rather than the bias term, and the lowest-frequency channel is 23.8 GHz in the present work compared with 10.65 GHz in Geer (2024a). These differences likely contribute to the observed discrepancies in the results.

Concerning the bias loss, $J_{bias}$ (Equation 6), in Experiment 1 (Figure 9a) it ranges between 0 and 1.68, whereas for Experiments 4 and 5 it is much lower, ranging between $10^{-6}$ and $10^{-5}$. By comparison, Geer (2024a) reports a bias loss ranging from 0 to less than 0.15.

As described in Section 2.1, the primary unknowns in this hybrid ML model—determined during training—are the evolving sea-ice concentration, the physical properties of the sea-ice and its snow cover, and a model to estimate the sea-ice surface emissivity from these properties. Together, these components allow the model to simulate the BTs presented in the previous sections. To illustrate this, maps of these components from Experiment 4 are presented. Note that this experiment is the first to include angular dependency and, compared with Experiment 5, reduces the mean differences between observed and simulated BTs for the first two channels by roughly 25–30%. Experiment 5 also included angular dependency but used $e_{\mathrm{bg}} = 0.6$ instead of $e_{\mathrm{bg}} = 0.7$.

Figure 10 shows the sea-ice concentration generated by the hybrid ML model on 1 April 2024, compared with ECMWF IFS data and their differences. At first glance, the ice extent is broadly similar in both cases, although ECMWF IFS identifies more ice around the Sea of Okhotsk. Notably, the hybrid ML model captures more detailed spatial structure, particularly in the western Arctic (coasts north of Greenland and Canada). Further investigations could explore whether the consistency with ECMWF IFS improves when considering a 2–3 day time delay, following the approach of Geer (2024a), or compare it with other sea-ice concentration products.

Figure 11 presents maps of the two empirical micro- and macro-physical properties on the model grid, along with the third property in observation space, from 1 to 3 April 2024. For properties 1 and 2, which are defined on the grid (the same grid used for sea-ice concentration), points where sea-ice concentration was lower than 0.2 were filtered out to focus mainly on sea-ice regions. Property 3, however, is defined in observation space; therefore, in this case, all property values at the observed locations for the specific days are shown.

The spatial patterns of properties 1 and 2 over these three days appear quite similar, although some differences can be detected. For example, on 2 April 2024, property 2 exhibits lower values than property 1 off the northern coast of Canada, while higher values are observed in the northern to slightly northeastern part of the Arctic Ocean. Property 3 varies over a smaller range than properties 1 and 2 during these days, mostly between -0.2 and 0.2, and displays spatial patterns that differ markedly from those of the other two properties. For instance, on 3 April 2024 (Figure 11, panels g, h and i), properties 1 and 2 show very high values around the Beaufort Sea and the southeastern

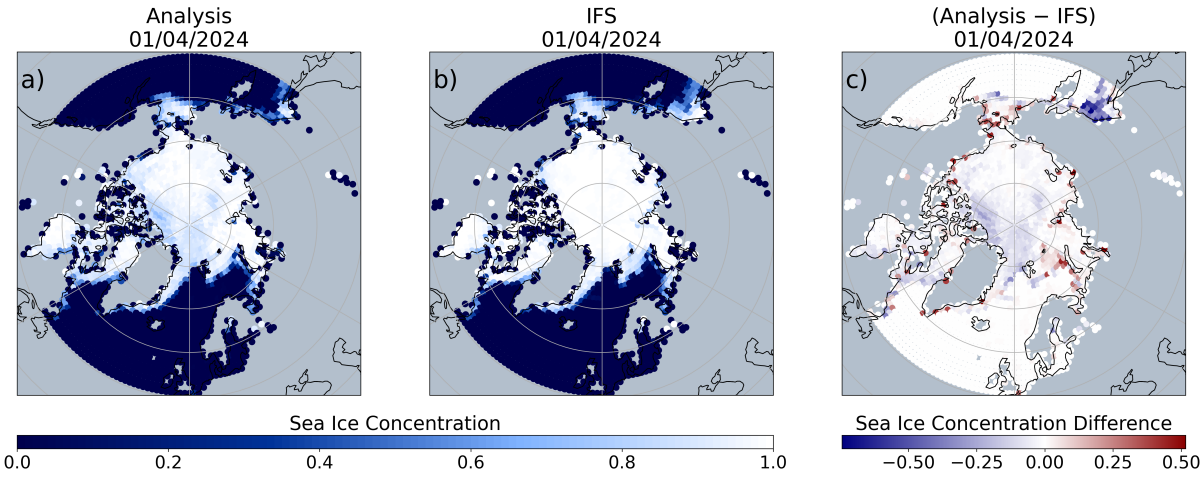


Figure 10: Sea-ice concentration map for 1 April 2024 from (a) the hybrid machine learning analysis (Experiment 4) and (b) the ECMWF IFS sea-ice initial conditions, together with (c) their difference (analysis minus IFS). The same colour scale is used for panels (a) and (b), with values ranging from 0 (no ice) to 1 (full ice) at each grid point.

Arctic Ocean, whereas property 3 exhibits negative values. Overall, all three properties exhibit day-to-day variability.

As noted previously, the empirical properties are learned along with the empirical sea-ice emissivity model and therefore do not have a precise or fixed physical interpretation. Nevertheless, the spatial patterns observed in properties 1 and 2 clearly reflect known characteristics of Arctic sea-ice, such as the distinctions between new ice and multi-year ice (Comiso, 2012; Lee et al., 2017).

Figure 12 shows the sea-ice emissivity for the same days across the different channels in observation space. Considerable variability in sea-ice emissivity is observed in areas where sea-ice concentration is below 0.2 (Figure 10a), which is expected due to the mixed contributions from ocean (low emissivity) and ice (higher emissivity) at the ice edge. In contrast, in regions dominated by sea-ice (concentrations above 0.5), the emissivity patterns are more homogeneous. Values lower than 0.9 are found in the southwestern Arctic Ocean, off the northern coasts of Greenland and Canada; these areas are characterized by multi-year ice. In contrast, ice emissivity values above 0.95 are observed in areas such as the Beaufort Sea, Chukchi Sea, Kara Sea, and Laptev Sea, where sea-ice is usually new. Differences between channels are difficult to identify, even though all plots use the same scale.

This motivated Figure 13, which illustrates two specific Arctic locations on 1 April 2024 (panel a): a multi-year ice site off the northern coast of Greenland and a new ice site in the Kara Sea, with the sea-ice emissivity across channels shown in panels b and c, respectively. At the first location, sea-ice emissivity is around 0.83 and decreases slightly with increasing channel frequency, whereas at the second location, it remains practically constant across channels. Therefore, the difference in sea-ice emissivity between these two locations exceeds 0.1.

These results are consistent with the findings of Gloersen et al. (1973), who showed that at a surface temperature of approximately 260 K, multi-year ice exhibits substantially

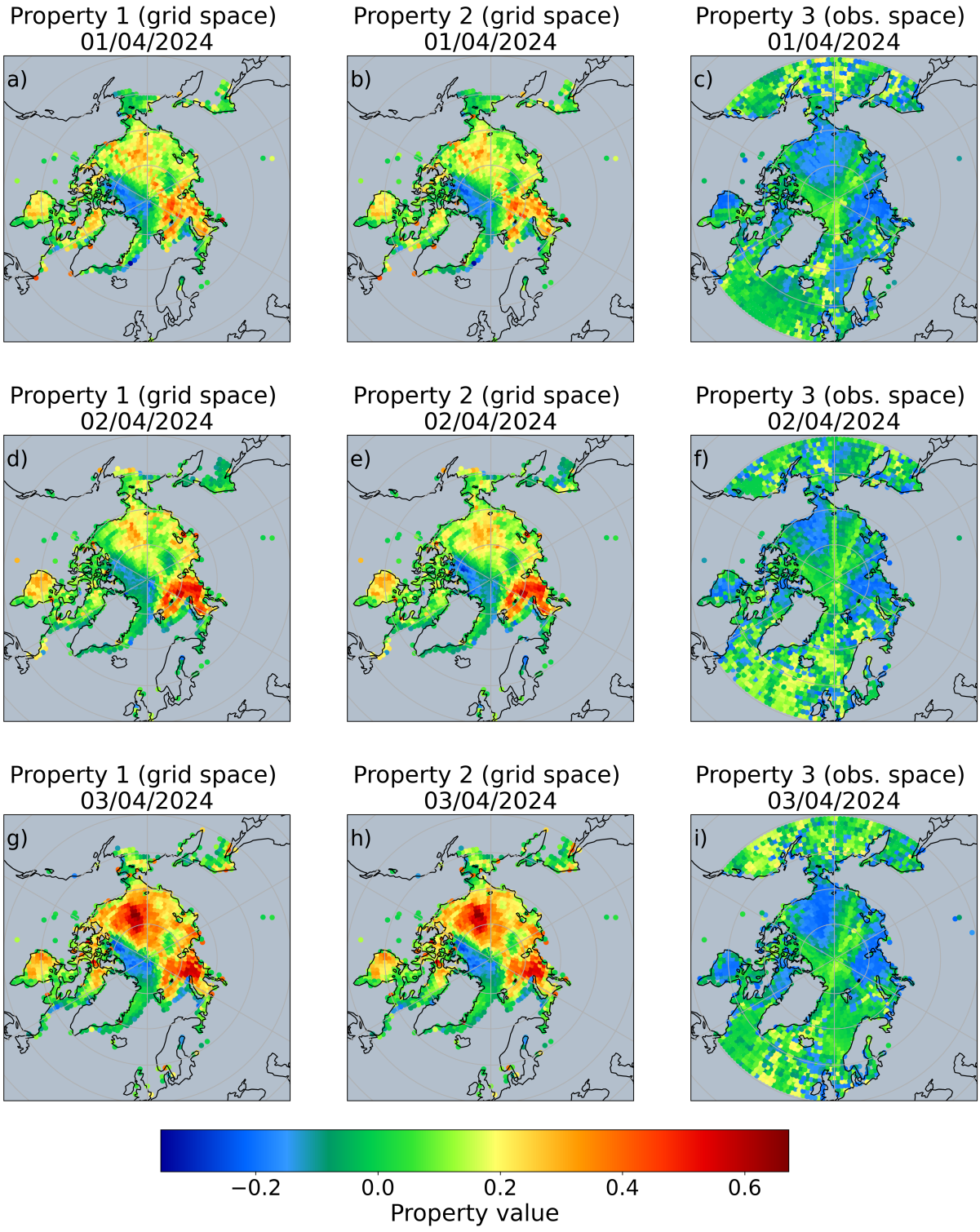


Figure 11: Maps of three empirical sea-ice properties for the Experiment 4 from 1–3 April 2024. Columns show Property 1 (grid space), Property 2 (grid space), and Property 3 (observation space), respectively. For grid-space properties, only points with sea-ice concentration $> 0.2$ are displayed. A single colorbar at the bottom is shared across all properties and days.

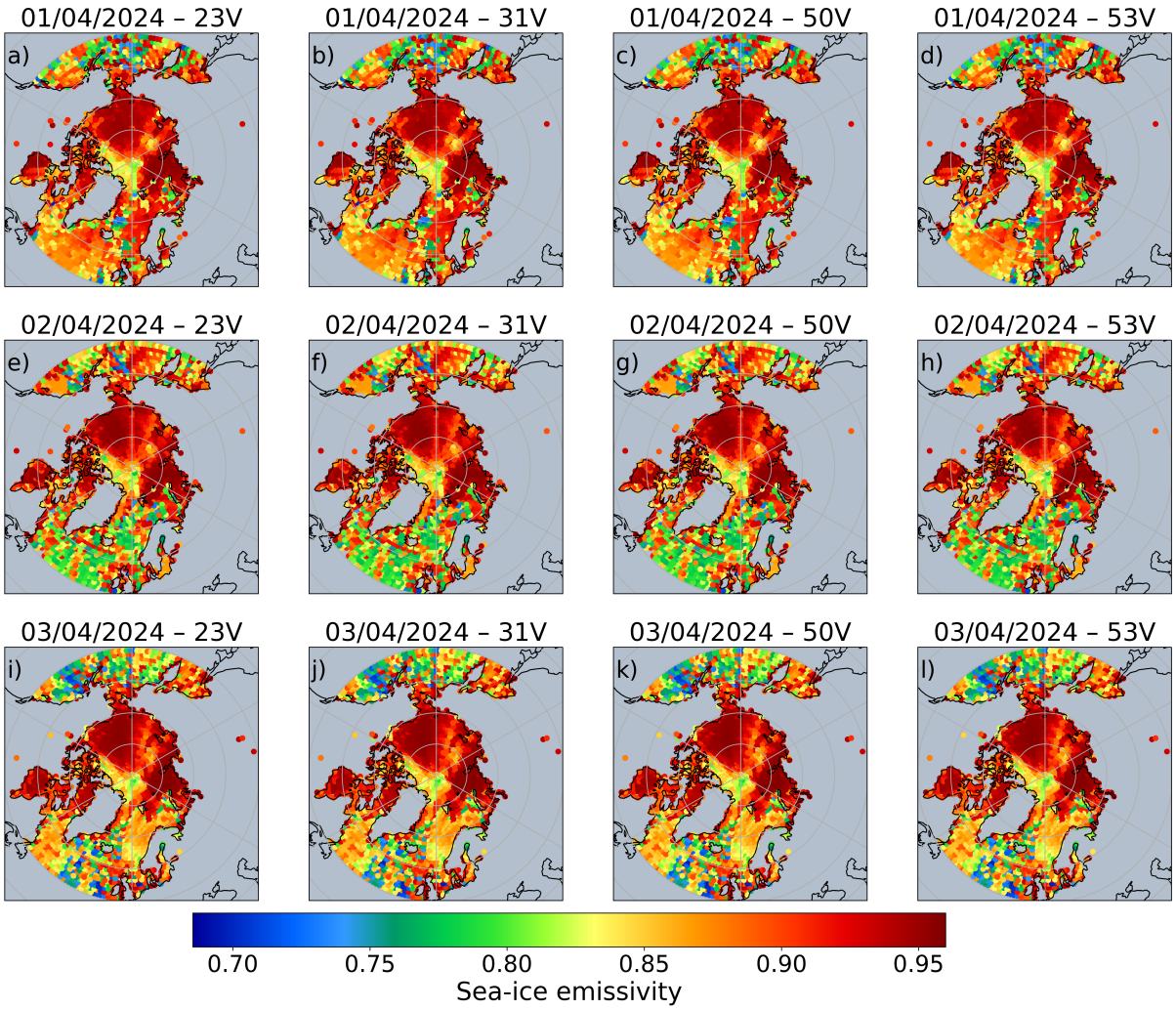


Figure 12: Maps of sea-ice emissivity from the Experiment 4 for 1–3 April 2024 (rows) for channels 23V, 31V, 50V, and 53V (columns). Colors indicate the sea-ice emissivity at each observation location. A single colorbar at the bottom is shared across all channels and days.

lower emissivity ($< 0.95$) than first-year or new ice ($\geq 0.95$) over wavelengths of roughly 0.8–11 cm (frequencies of approximately 3–37 GHz). Moreover, their study demonstrated a monotonic decrease in multi-year ice emissivity with increasing frequency, from about 0.95 at 3 GHz (11 cm) to approximately 0.8 at 37 GHz (0.8 cm). They attributed this behaviour primarily to enhanced volume scattering within multi-year ice, rather than to differences in surface roughness or salinity.

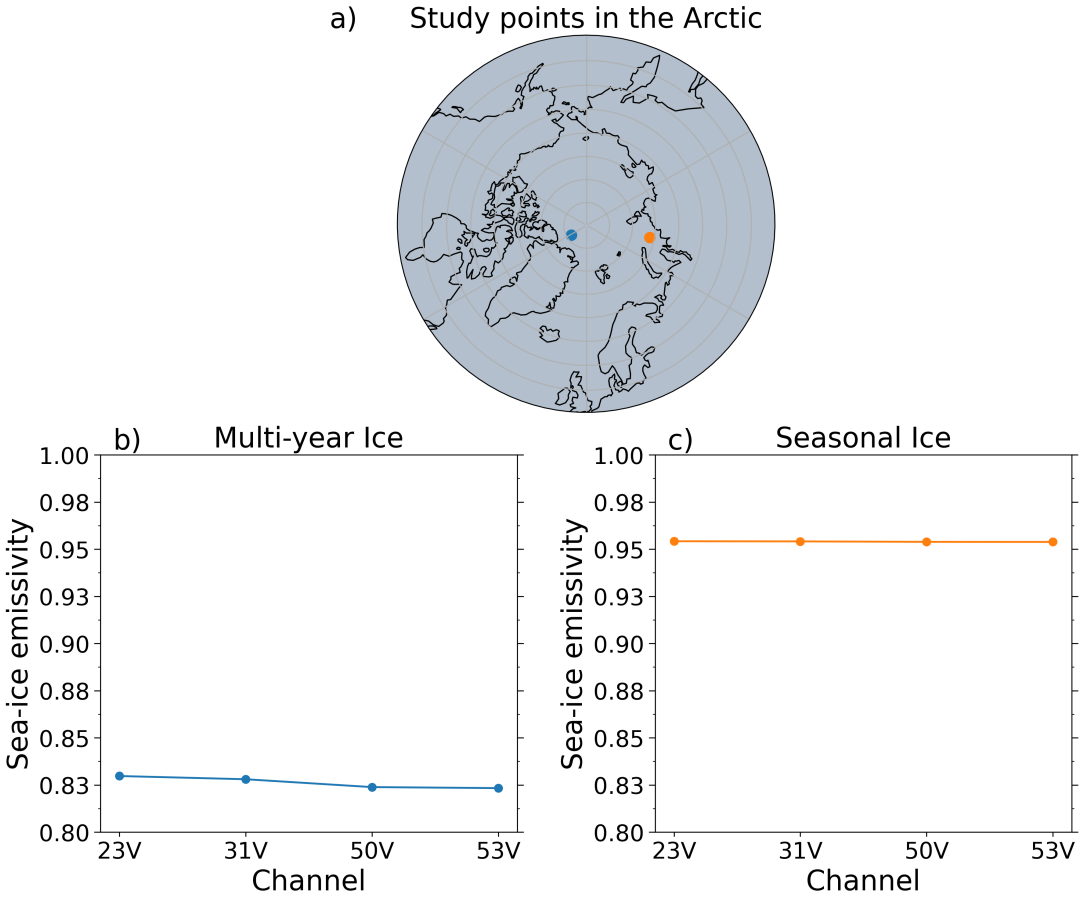


Figure 13: (a) Map showing a multi-year ice (blue) and a new ice (orange) study point. (b, c) Sea-ice emissivity across channels 23V, 31V, 50V, and 53V at the multi-year and new ice locations, respectively, on 1st April 2024, derived from the Experiment 4.

# 4 Conclusions

This report summarises the first year of work of my EUMETSAT fellowship, which focused on the development, training, and evaluation of a hybrid empirical–physical ML model applied for the first time to AMSU-A observations over sea-ice surfaces. This application to AMSU-A represents the main novelty of the study and lays the foundation for application to other cross-track sensors.

The work began with the retrieval of satellite observations and corresponding ECMWF IFS outputs from the ECMWF MARS archive in .odb and .grib formats. Particular attention was devoted to dataset construction, including data cleaning, quality control, consistency checks between satellite observations and auxiliary fields, and transformation into the format required by the hybrid model.

A substantial effort was then dedicated to understanding the model architecture, including the interaction between its physical and empirical components, in order to adapt it to the specific characteristics of AMSU-A. This process included the implementation of viewing-angle dependency required for cross-track sensors and was conceived with future extensions in mind, particularly the potential simultaneous integration of observations from multiple satellite platforms, including both cross-track and conical-scanning instruments.

Once the dataset was fully prepared and the model adapted to ingest AMSU-A observations, several experiments were conducted to iteratively adjust the model parameters, aiming to optimise its performance according to the AMSU-A instrument specifications.

The first major objective (achieved prior to the EUMETSAT Conference 2025) was to successfully run the model with AMSU-A inputs and obtain physically realistic outputs. In this respect, the initial performance was encouraging. Even without AMSU-A–specific tuning or explicit scan-angle dependency, the model already reproduced the observed BT patterns and absolute values reasonably for the two lowest-frequency channels (23V and 31V), which are the most sensitive to surface conditions. For these channels, the mean Obs−Sim differences were substantially reduced relative to the ECMWF IFS FG departures to observations, with approximate improvements of 99% and 77% for the 23V and 31V channels, respectively. At the higher-frequency channels, however, biases in this first run were larger and some spatial patterns were not fully captured. While this may partly reflect atmospheric contamination, as these channels are more affected by atmospheric absorption, it also highlights the need for further refinement of the model settings to improve agreement with the observations.

The second objective therefore consisted of the incremental optimisation of the model parameters, carried out through a series of controlled experiments aimed at progressively reducing the biases relative to AMSU-A observations. To illustrate, Experiment 1 used the parameter configuration largely based on Geer (2024a), which was originally tuned for AMSR2, whereas in Experiment 3 the parameters were tuned specifically for AMSU-A. Considering all scan positions, this tuning led to dramatic reductions in mean Obs−Sim differences for the higher-frequency channels by roughly 91% for 31V, 95% for 50V, and 99% for 53V, while the 23V channel, already well reproduced in the first experiment, remained largely unchanged. These results confirm that parameter tuning alone is a key component in improving the model's performance.

The third objective addressed the implementation of scan-angle dependency. When considering large viewing angles, its inclusion led to a modest but systematic improvement, with reductions in mean Obs−Sim differences of approximately 60% for the first channel and 17% for the third channel. In addition, both the standard deviation and the RMSE decrease by about 1–4% across all four channels, indicating reduced error variability and more consistent model behaviour at large scan angles. By contrast, for observations near nadir, its impact is more mixed, with slightly increased mean Obs−Sim differences but small reductions in standard deviation and RMSE. This behaviour warrants further investigation to better understand its physical origin and to identify possible ways to improve the current implementation.

It is important to note that the hybrid ML model currently assumes a specular reflec-

tion behaviour, with the effective emissivity seen by the instrument obtained through a deterministic, scan-angle-dependent polarization mixing based on the known instrument geometry. As a starting point, the model ignores variations of emissivity with the zenith angle, which is adequate in many situations where surface scattering is not dominated by boundary reflections or strongly oriented snow and ice particles. A natural extension of this work would therefore be to include the zenith angle as an additional predictor in the `SeaiceEmisNN` layer, enabling the model to account for such angular dependence and potentially improve the accuracy of simulated radiances and retrievals in specific regions or surface type regimes.

While most of the analysis in this report has focused on differences between observed and simulated BTs in order to identify the most suitable model configuration, the hybrid model also produces a range of additional outputs. These include daily maps of the sea-ice and its properties, as well as a new empirical model for the sea-ice surface emissivity. A first, preliminary analysis of some of these outputs, mainly for selected day (1 April 2024) from experiment 4 has been discussed. On this day, the daily sea-ice concentration fields generated by the hybrid model exhibit more detailed spatial structure. The empirical–physical properties show distinct spatial patterns: the two properties defined in grid space appear broadly similar, whereas the property defined in observation space exhibits a reduced range of variability and markedly different spatial distributions. All properties also show clear day-to-day variability (1–3 April). The retrieved sea-ice emissivity at two selected locations—representative of multi-year ice and newly formed ice—agrees well with previous findings in the literature, with systematically lower emissivity values over multi-year ice.

The loss terms used during training were only briefly discussed in this report. A more detailed analysis of their relative contributions could be beneficial, either to introduce additional constraints that improve model performance or to simplify the cost function if certain terms prove to have limited impact. In the current study, the hybrid model used a simple neural network with two layers comprising 7 and 2 neurons, respectively, trained for 8 epochs with a batch size of 1024. Additional sensitivity experiments exploring these hyperparameters could help optimise the model configuration further. From a technical perspective, an important practical lesson from this work is the necessity of ensuring full reproducibility. Achieving consistent and repeatable results required using a controlled software environment, including a specific Python version, enabling deterministic TensorFlow operations, and setting a fixed random seed in the code.

So far, this study has been restricted to AMSU-A onboard METOP-B, the four lowest-frequency channels, and the Arctic region. The training dataset, however, already includes AMSU-A observations from METOP-C, NOAA-15, NOAA-18, and NOAA-19. To fully exploit these additional sensors, some adjustments to the hybrid model would be required. The dataset could also be extended to cover a longer time period, which would be particularly relevant for METOP-B and METOP-C. For the NOAA sensors, the available period is limited by their retirement dates: NOAA-15 on 19 August 2025, NOAA-18 on 6 June 2025, and NOAA-19 on 13 August 2025. Exploring this extended dataset could provide additional insights and benefits for the model.

Finally, the current implementation of the model is based on TensorFlow and has proven to be robust and efficient. Nevertheless, migrating the framework to PyTorch could be

considered in the future, given its widespread adoption within the research community and potential advantages in terms of flexibility and long-term maintainability. Regarding applications of this study, while the original idea of deploying this framework within limited-area models such as the Applications of Research to Operations at MEsoscale (AROME)-Arctic forecast or CARRA-2 reanalysis systems now appears less feasible (in the context of this fellowship) in view of the change in scientific direction, it may still be possible—subject to support from ECMWF—to assess its impact through a dedicated DA experiment within the ECMWF system.

# 5 Personal reflections

Before outlining the next steps of my EUMETSAT fellowship in Section 6, I think it is worth briefly reflecting on its broader context and on what these initial months have represented for me.

This first year has been largely dedicated to training and skill development and, as such, has not been without challenges. While several competencies acquired during my PhD proved transferable, the scientific focus of this fellowship differs substantially from my previous research. Consequently, this year marked my first extensive experience working with satellite observations, interpreting their geophysical signatures, conducting experiments in a high-performance computing environment, and developing a deeper understanding of NWP and DA. In addition, I have acquired a solid grounding in ML and applied several of its techniques.

Although the learning curve has sometimes been steep and, at times, overwhelming, the experience has been highly rewarding. I have thoroughly enjoyed working on this topic, which I find both challenging and motivating. Overall, I am very satisfied with the progress achieved and look forward to continuing this journey in the months ahead. The skills and knowledge I have acquired during this first year provide a solid foundation for the remainder of the fellowship, where I will be redirecting the research focus, as described in Section 6.

# 6 Next steps

While the work presented here will continue in the short term, and I will make every effort to further develop it toward a potential publication, it is important to acknowledge that this will require additional work and may prove challenging in light of the ongoing transition to a new research direction following the departure of my primary advisor.

Over the past months, a revised research plan for the rest of my fellowship has been developed, ensuring alignment with the overarching goals of the EUMETSAT Joint Research Fellowship and the strategic interests of both DMI and Norwegian Meteorological Institute (MetNO). Building on the knowledge and experience gained over the past year, the revised proposal remains grounded in satellite-based Earth observation. It continues to focus on complex and remote polar environments, with particular attention to fast ice, a sea-ice type of high climatic and socio-economic relevance (George et al., 2004; Freitas et al., 2012; Dammann et al., 2018; Fraser et al., 2023) which, although subject to some recent study, remains underrepresented in operational contexts (Zhao et al., 2020; Wulf

et al., 2024; Huang et al., 2024; Dammann et al., 2025). Furthermore, given the growing recognition within the scientific and operational communities—including meteorologists, glaciologists, and polar observation agencies—of the potential for ML to complement, and in some cases even replace, traditional DA in certain applications, this proposal will adopt fully ML-based approaches.

In particular, the objective of this study is the automated detection and operational monitoring of fast ice in polar regions, using Synthetic Aperture Radar (SAR) data (e.g., Sentinel-1). As an active microwave sensor, SAR emits its own signal and measures the backscattered response from the surface, enabling the retrieval of fine-scale ice structures at high spatial resolution. Its capacity to deliver all-weather, day-and-night observations makes SAR especially valuable for both operational and research applications in polar regions (Karvonen, 2017; Soldal et al., 2019; Dammann et al., 2019), where optical data are often severely limited by persistent cloud cover and prolonged periods of polar darkness. Nevertheless, the use of SAR for pan-regional coverage and consistent multi-year monitoring remains constrained by limited swath width and spatially and temporally uneven image acquisition, underscoring both the substantial potential and the current limitations of SAR-based approaches in operational contexts.

The Automated Sea-Ice Products (ASIP) system developed at DMI (Wulf et al., 2024, 2025) already provides high-resolution sea-ice concentration fields with calibrated uncertainties from Sentinel-1 SAR and AMSR2 across all Arctic seasons. However, ASIP does not currently include explicit information on fast ice. The proposed work aims to fill this important gap by adding an operationally relevant fast ice component to an already well-established framework. Initial efforts will focus on the period 2018–2021, for which a well-curated dataset of colocated Sentinel-1A/B SAR imagery and ice charts is already available. The study will concentrate on the Arctic, with particular emphasis on regions around Greenland and the Canadian Arctic Archipelago, while retaining the potential for future extension to Antarctica.

The planned workflow over the coming months consists of: (i) identifying days and regions characterized by fast ice based on operational ice charts; (ii) collocating SAR imagery covering these areas, including images from preceding days to capture temporal persistence; and (iii) assembling these data into a dedicated ML training dataset. This dataset will then be used to train a modified U-Net architecture (Ronneberger et al., 2015) tailored to fast ice detection.

This work could enable in the future the production of automated ice charts with explicit fast ice information, support the integration of these data into numerical weather prediction systems such as AROME-Arctic, contribute to future sea-ice reanalysis products, and enhance sea-ice forecasting and prediction capabilities.

# 7 Acknowledgments

Cristina is supported by the EUMETSAT Fellowship Programme. She deeply acknowledges the guidance and support of her supervisors, Fabrizio Baordo and Stéphanie Guedj, throughout this first year. Their mentorship has been instrumental in helping her navigate this new research path in satellite data, DA, and ML, from which she has learned a great deal. She also thanks Alan Geer for his time and valuable assistance, particu-

larly when addressing challenges related to the hybrid ML model. She is also grateful to Suman Singha, Tore Wulf, Cyril Palerme, and Matilde Brandt Kreiner for their support in reshaping the fellowship. Finally, she warmly acknowledges Rune Carbuhn and Julia Sommer for welcoming her into their research group upon her arrival at DMI and for their efforts to integrate her into the group..

# References

Baordo, F. and Geer, A. J. (2016). Assimilation of SSMIS humidity-sounding channels in all-sky conditions over land using a dynamic emissivity retrieval. *Quarterly Journal of the Royal Meteorological Society*, 142(700):2854–2866. DOI: https://doi.org/10.1002/qj.2873.

Bauer, P., Moreau, E., Chevallier, F., and O'Keeffe, U. (2006). Multiple-scattering microwave radiative transfer for data assimilation applications. *Quarterly Journal of the Royal Meteorological Society*, 132(617):1259–1281. DOI: https://doi.org/10.1256/qj.05.153.

Bormann, N. (2022). Accounting for lambertian reflection in the assimilation of microwave sounding radiances over snow and sea-ice. *Quarterly Journal of the Royal Meteorological Society*, 148(747):2796–2813. DOI: https://doi.org/10.1002/qj.4337.

Buehner, M., Caya, A., Pogson, L., Carrieres, T., and Pestieau, P. (2013). A new environment canada regional ice analysis system. *Atmosphere-Ocean*, 51(1):18–34. DOI: https://doi.org/10.1080/07055900.2013.772927.

Comiso, J. C. (2012). Large decadal decline of the arctic multiyear ice cover. *Journal of Climate*, 25(4):1176–1193. DOI: https://doi.org/10.1175/JCLI-D-11-00113.1.

Dammann, D. O., Eicken, H., Mahoney, A. R., Meyer, F. J., and Betcher, S. (2018). Assessing sea ice trafficability in a changing arctic. *Arctic*, 71(1):59–75. DOI: https://www.jstor.org/stable/26387330.

Dammann, D. O., Eriksson, L. E. B., Mahoney, A. R., Eicken, H., and Meyer, F. J. (2019). Mapping pan-arctic landfast sea ice stability using sentinel-1 interferometry. *The Cryosphere*, 13(2):557–577. DOI: https://doi.org/10.5194/tc-13-557-2019.

Dammann, D. O., Vöge, M. L., Hughes, B., Haas, C., and Body, N. (2025). Iceview: Operational landfast sea ice mapping with insar. Poster presented at the 26th International Ice Charting Working Group (IICWG) Meeting. 22–26 September 2025. Available online: https://o365coloradoedu.sharepoint.com/sites/NSIDC-MT-IICWG/Shared%20Documents/Forms/AllItems.aspx?viewid=ccaaaac9%2D960b%2D4742%2Db2b3%2Dfad8296659fd&id=%2Fsites%2FNSIDC%2DMT%2DIICWG%2FShared%20Documents%2FGeneral%2FIICWG%2DXXVI%2D2025%2FIICWG%2DXXVI%2DPosters%2FIICWGXXVI%5FDammann%5FPoster%2Epdf&parent=%2Fsites%2FNSIDC%2DMT%2DIICWG%2FShared%20Documents%2FGeneral%2FIICWG%2DXXVI%2D2025%2FIICWG%2DXXVI%2DPosters.

Fellous, J.-L. (2008). Towards a global climate observing system. *Interdisciplinary Science Reviews*, 33(1):83–94. DOI: https://doi.org/10.1179/030801808X259952.

Fraser, A. D., Wongpan, P., Langhorne, P. J., Klekociuk, A. R., Kusahara, K., Lannuzel, D., Massom, R. A., Meiners, K. M., Swadling, K. M., Atwater, D. P., et al. (2023). Antarctic landfast sea ice: A review of its physics, biogeochemistry and ecology. *Reviews of Geophysics*, 61(2). DOI: https://doi.org/10.1029/2022RG000770.

Freitas, C., Kovacs, K. M., Andersen, M., Aars, J., Sandven, S., Skern-Mauritzen, M., Pavlova, O., and Lydersen, C. (2012). Importance of fast ice and glacier fronts for female polar bears and their cubs during spring in svalbard, norway. *Marine Ecology Progress Series*, 447:289–304. DOI: https://doi.org/10.3354/meps09516.

Geer, A. J. (2024a). Joint estimation of sea ice and atmospheric state from microwave imagers in operational weather forecasting. *Quarterly Journal of the Royal Meteorological Society*, 150(763):3796–3826. DOI: https://doi.org/10.1002/qj.4797.

Geer, A. J. (2024b). Simultaneous inference of sea ice state and surface emissivity model using machine learning and data assimilation. *Journal of Advances in Modeling Earth Systems*, 16(7). DOI: https://doi.org/10.1029/2023MS004080.

Geer, A. J., Lonitz, K., Weston, P., Kazumori, M., Okamoto, K., Zhu, Y., Liu, E. H., Collard, A., Bell, W., Migliorini, S., et al. (2018). All-sky satellite data assimilation at operational weather forecasting centres. *Quarterly Journal of the Royal Meteorological Society*, 144(713):1191–1217. DOI: https://doi.org/10.1002/qj.3202.

George, J. C., Huntington, H. P., Brewster, K., Eicken, H., Norton, D. W., and Glenn, R. (2004). Observations on shorefast ice dynamics in arctic alaska and the responses of the iñupiat hunting community. *Arctic*, 57(4):363–374. DOI: https://www.jstor.org/stable/40512640.

Gloersen, P., Nordberg, W., Schmugge, T. J., Wilheit, T. T., and Campbell, W. J. (1973). Microwave signatures of first-year and multiyear sea ice. *Journal of Geophysical Research*, 78(18):3564–3572. DOI: https://doi.org/10.1029/JC078i018p03564.

González-Flórez, C., Baordo, F., and Guedj, S. (2025). Enhancing microwave satellite data assimilation over sea ice with hybrid machine learning modelling. Poster presented at the EUMETSAT Conference 2025. 15–19 September 2025. Available online: https://imagine.eumetsat.int/smartViews/view?view=EMSC.

Huang, Y., Ren, Y., and Li, X. (2024). Deep learning techniques for enhanced sea-ice types classification in the beaufort sea via SAR imagery. *Remote Sensing of Environment*, 308. DOI: https://doi.org/10.1016/j.rse.2024.114204.

Inoue, J. (2021). Review of forecast skills for weather and sea ice in supporting arctic navigation. *Polar Science*, 27. DOI: https://doi.org/10.1016/j.polar.2021.100523.

Karvonen, J. (2017). Baltic sea ice concentration estimation using sentinel-1 sar and amsr2 microwave radiometer data. *IEEE Transactions on Geoscience and Remote Sensing*, 55(5):2871–2883. DOI: https://ieeexplore.ieee.org/document/7857748.

Kilic, L., Prigent, C., Jimenez, C., Turner, E., Hocking, J., English, S., Meissner, T., and Dinnat, E. (2023). Development of the SURface Fast Emissivity Model for Ocean (SURFEM-Ocean) based on the PARMIO radiative transfer model. *Earth and Space*

*Science*, 10(4):e2022EA002785. DOI: https://doi.org/10.1029/2022EA002785.

Kingma, D. P. and Ba, J. (2015). A method for stochastic optimization. In *International Conference on Learning Representations (ICLR) 2015, Poster Presentations*, pages 1–15. ICLR / arXiv. DOI: https://arxiv.org/abs/1412.6980v5.

Lee, S.-M., Sohn, B.-J., and Kim, S.-J. (2017). Differentiating between first-year and multiyear sea ice in the arctic using microwave-retrieved ice emissivities. *Journal of Geophysical Research: Atmospheres*, 122(10):5097–5112. DOI: https://doi.org/10.1002/2016JD026275.

Madec, G. and the NEMO System Team (2023). *NEMO Ocean Engine Reference Manual*. DOI: https://zenodo.org/records/14515373.

Naakka, T., Köhler, D., Nordling, K., Räisänen, P., Lund, M. T., Makkonen, R., Merikanto, J., Samset, B. H., Sinclair, V. A., Thomas, J. L., et al. (2025). Polar winter climate change: strong local effects from sea ice loss, widespread consequences from warming seas. *EGUsphere*, 25(14):8127–8145. DOI:https://doi.org/10.5194/acp-25-8127-2025.

Ronneberger, O., Fischer, P., and Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*. DOI: https://doi.org/10.48550/arXiv.1505.04597.

Rousset, C., Vancoppenolle, M., Madec, G., Fichefet, T., Flavoni, S., Barthélemy, A., Benshila, R., Chanut, J., Levy, C., Masson, S., and Vivier, F. (2015). The Louvain-La-Neuve sea ice model LIM3.6: global and regional capabilities. *Geoscientific Model Development*, 8(10):2991–3005. DOI: https://doi.org/10.5194/gmd-8-2991-2015.

Saunders, R., Hocking, J., Turner, E., Rayer, P., Rundle, D., Brunel, P., Vidot, J., Roquet, P., Matricardi, M., Geer, A. J., Bormann, N., and Lupu, C. (2018). An update on the RTTOV fast radiative transfer model (currently at version 12). *Geoscientific Model Development*, 11(7):2717–2737. DOI: https://doi.org/10.5194/gmd-11-2717-2018.

Soldal, I. H., Dierking, W., Korosov, A., and Marino, A. (2019). Automatic detection of small icebergs in fast ice using satellite wide-swath SAR images. *Remote Sensing*, 11(7):806. DOI: https://doi.org/10.3390/rs11070806.

Turner, J. and Marshall, G. J. (2011). *Climate change in the polar regions*. Cambridge University Press. DOI: https://doi.org/10.1017/CBO9780511975431.

Wang, Z. Q. and Randriamampianina, R. (2021). The impact of assimilating satellite radiance observations in the copernicus european regional reanalysis (CERRA). *Remote Sensing*, 13(3):426. DOI: https://doi.org/10.3390/rs13030426.

Wulf, T., Buus-Hinkler, J., Singha, S., Dasgupta, N., Athanasiadis, A., and Kreiner, M. B. (2025). A decade of sea ice concentration retrievals from sentinel-1. *SSRN*. DOI: http://dx.doi.org/10.2139/ssrn.5222853.

Wulf, T., Buus-Hinkler, J., Singha, S., Shi, H., and Kreiner, M. B. (2024). Pan-arctic sea

ice concentration from SAR and passive microwave. *The Cryosphere*, 18(11):5277–5300. DOI: https://doi.org/10.5194/tc-18-5277-2024.

Yang, J., Gong, P., Fu, R., Zhang, M., Chen, J., Liang, S., Xu, B., Shi, J., and Dickinson, R. (2013). The role of satellite remote sensing in climate change studies. *Nature Climate Change*, 3:875–883. DOI: https://doi.org/10.1038/nclimate1908.

Zhao, J., Cheng, B., Vihma, T., Heil, P., Hui, F., Shu, Q., Zhang, L., and Yang, Q. (2020). Fast Ice Prediction System (FIPS) for land-fast sea ice at Prydz Bay, East Antarctica: an operational service for CHINARE. *Annals of Glaciology*, 61(83):271–283. DOI: https://doi.org/10.1017/aog.2020.46.